Technical Report 1257

Validating Future Force Performance Measures (Army Class): End of Training Longitudinal Validation

Deirdre J. Knapp (Ed.)
Tonia S. Heffner (Ed.)

September 2009

United States Army Research Institute
for the Behavioral and Social Sciences

U.S. Army Research Institute
for the Behavioral and Social Sciences

A Directorate of the Department of the Army
Deputy Chief of Staff, G1

Authorized and approved for distribution:

MICHELLE SAMS, Ph.D.
Director

Research accomplished under contract
for the Department of the Army

Human Resources Research Organization

Technical review by

J. Douglas Dressel, U.S. Army Research Institute
Trueman R. Tremble, U.S. Army Research Institute

NOTICES

DISTRIBUTION: Primary distribution of this Technical Report has been made by ARI. Please address correspondence concerning distribution of reports to: U.S. Army Research Institute for the Behavioral and Social Sciences, Attn: DAPE-ARI-ZXM, 2511 Jefferson Davis Highway, Arlington, Virginia 22202-3926.

FINAL DISPOSITION: This Technical Report may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The findings in this Technical Report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

REPORT DOCUMENTATION PAGE

1. REPORT DATE (dd-mm-yy)	2. REPORT TYPE	3. DATES COVERED (from. . . to)
September 2009	Final Report	January 2008 – December 2008

4. TITLE AND SUBTITLE	5a. CONTRACT OR GRANT NUMBER
Validating Future Force Performance Measures (Army Class): End of Training Longitudinal Validation	DASW01-03-D-0015, DO #0029
	5b. PROGRAM ELEMENT NUMBER 622785
6. AUTHOR(S) Knapp, Deirdre J. & Heffner, Tonia S. (Editors)	5c. PROJECT NUMBER A790
	5d. TASK NUMBER 257
	5e. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
Human Resources Research Organization 66 Canal Center Plaza, Suite 700 Alexandria, Virginia 22314	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences ATTN: DAPE-ARI-RS 2511 Jefferson Davis Highway Arlington, VA 22202-3926	10. MONITOR ACRONYM ARI 11. MONITOR REPORT NUMBER Technical Report 1257

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

Contracting Officer's Representative and Subject Matter POC: Dr. Tonia Heffner

14. ABSTRACT *(Maximum 200 words)*:

The Army needs the best personnel to meet the emerging demands of the 21st century. Accordingly, the Army is seeking recommendations on new experimental predictor measures that could enhance entry-level Soldier selection and classification decisions, in particular, measures of non-cognitive attributes (e.g., interests, values, temperament). The U. S. Army Research Institute for the Behavioral and Social Sciences (ARI) is conducting a longitudinal criterion-related validation research effort to collect data to inform these recommendations.

Data on experimental predictors were collected from about 11,000 Soldiers. Training criterion data were collected for differing subsets of the predictor sample in the first of three planned criterion measurement points. Soldiers were drawn from two samples: (a) job-specific samples targeting six entry-level Military Occupational Specialties (MOS) and (b) an Army-wide sample with no MOS-specific requirements. In the analyses reported here, the value of the experimental predictor measures to enhance new Soldier selection was examined. Overall, many of the experimental predictors significantly incremented the Armed Forces Qualification Test (AFQT) in predicting Soldier performance and retention during training. In addition, the experimental predictors generally exhibited smaller subgroup mean differences (by gender, race, and ethnicity) than the AFQT.

15. SUBJECT TERMS

Behavioral and social science Personnel Criterion-related validation Selection and classification Manpower

SECURITY CLASSIFICATION OF			19. LIMITATION OF ABSTRACT	20. NUMBER OF PAGES	21. RESPONSIBLE PERSON
16. REPORT	17. ABSTRACT	18. THIS PAGE			Ellen Kinzer Technical Publications Specialist (703) 602-8049
Unclassified	Unclassified	Unclassified	Unlimited	83	

Standard Form 298

Technical Report 1257

Validating Future Force Performance Measures (Army Class): End of Training Longitudinal Validation

Deirdre J. Knapp (Ed.)
Tonia S. Heffner (Ed.)

Personnel Assessment Research Unit
Michael G. Rumsey, Chief

U.S. Army Research Institute for the Behavioral and Social Sciences
2511 Jefferson Davis Highway, Arlington, Virginia 22202-3926

September 2009

Army Project Number
622785 A790

Personnel, Performance
and Training Technology

ACKNOWLEDGEMENTS

There are a large number of individuals not listed as authors who have contributed significantly to the work described in this report. Drs. Kimberly Owens and Richard Hoffman of the U.S. Army Research Institute for Behavioral and Social Sciences (ARI) provided oversight and support during the training criterion development and data collection efforts. The Human Resources Research Organization (HumRRO) personnel primarily responsible for development of the training criterion measures included Drs. Karen Moriarty, Teresa Russell, Patricia Keenan, Gordon Waugh, Laura Ford, Kevin Bradley, and Mr. Roy Campbell. Data collection support was provided by a number of individuals from both ARI and HumRRO, including those listed below:

ARI: Nehama Babin, Elizabeth Brady, Doug Dressel, Kelly Ervin, Tonia Heffner, Ryan Hendricks, Rich Hoffman, Colanda Howard, Arwen Hunter, Kimberly Owens, Peter Schaefer, Teresa Taylor, Mike Wesolak, Len White, and Mark Young

HumRRO: Matthew Allen, Joe Caramagno, John Fisher, Patricia Keenan, Julisara Mathew, Alicia Sawyer, Jim Takitch, Shonna Waters, and Elise Weaver

Drasgow Consulting Group: Gabriel Lopez

Dr. Karen Moriarty (HumRRO) and Ms. Sharon Meyers (ARI) prepared the training measures for computer-based administration. Ms. Ani DiFazio was responsible for preparing the analysis database, with data cleaning and scoring assistance from several people already listed as well as Dr. Matthew Trippe, Ms. Dalia Diab (HumRRO), and Dr. Arwen Hunter (ARI). Dr. Dan Putka (HumRRO) provided statistical consultation and advice.

We are, of course, also indebted to the military and civilian personnel who supported our test development and data collection efforts, particularly those Soldiers and noncommissioned officers (NCOs) who participated in the research.

VALIDATING FUTURE FORCE PERFORMANCE MEASURES (ARMY CLASS): END OF TRAINING LONGITUDINAL VALIDATION

EXECUTIVE SUMMARY

Research Requirement:

The Army needs the best personnel to meet the emerging demands of the 21st century. Selecting and classifying these Soldiers requires new predictor measures that assess attributes not currently covered by the existing Armed Forces Qualification Test (AFQT), in particular measures of non-cognitive attributes (e.g., interests, values, and temperament). One of the objectives of the "Army Class" research program is to provide the Army with recommendations on which new experimental predictor measures evidence the greatest potential to enhance new Soldier selection and classification. The present report documents the first stages of a longitudinal criterion-related validation research effort conducted to advance this objective.

Procedure:

Predictor data were collected from about 11,000 entry-level enlisted Soldiers representing all Components (Regular Army, Reserve, National Guard). Criterion data were collected at the end of training. Soldiers were drawn from two samples: (a) job-specific samples targeting six entry-level Military Occupational Specialties (MOS) and (b) an Army-wide sample with no MOS-specific requirements. The experimental predictors were administered to new Soldiers as they entered the Army through one of four reception battalions. The predictor measures included (a) three temperament measures (Assessment of Individual Motivation [AIM], Tailored Adaptive Personality Assessment System (TAPAS), and Rational Biodata Inventory [RBI]), (b) a predictor situational judgment test (PSJT), and (c) two person-environment (P-E) fit measures (Work Preferences Assessment [WPA] and Army Knowledge Assessment [AKA]). In addition, we also obtained scores through administrative records on the Assembling Objects (AO) test, a spatial ability measure currently administered with the Armed Services Vocational Aptitude Battery (ASVAB). Two predictor measures (AIM and TAPAS) were added to the research to support a short-term requirement to identify predictors that could immediately be put into operational use by the Army (i.e., the *Expanded Enlistment Eligibility Metrics* [EEEM] initiative).

The criterion measures were administered to Soldiers in the six job-specific samples at the end of training. The criterion measures administered were (a) an MOS-specific job knowledge test (JKTs), (b) MOS-specific and Army-wide performance ratings collected from training instructors and peers, and (c) a questionnaire measuring Soldiers' experiences and attitudes towards the Army through training (the Army Life Questionnaire [ALQ]). For all Regular Army Soldiers, we obtained data on attrition (through the first 6 months of service) and for all Soldiers, we obtained data on performance during training from administrative records.

Two series of analyses were conducted. The first consisted of estimating and analyzing the incremental validity of the experimental predictors over the existing AFQT, across multiple performance and retention-related criteria. The second series of analyses involved estimating the

subgroup differences on the experimental predictor measures (by gender, race, and ethnicity) and comparing them to those observed for the existing AFQT.

Findings:

In regards to the incremental validity analyses, the experimental predictors consistently demonstrated the potential to significantly increment the AFQT in predicting both performance and retention-related criteria, including 6-month attrition. On the performance-related criteria, the experimental predictors yielded incremental validity estimates (ΔRs) that ranged from .01 to upwards of .35, on the more behaviorally-based criteria (a 648% gain in R over the AFQT). Among the experimental predictors, the RBI, the TAPAS, and the AIM, followed by the WPA, generally evidenced the greatest potential for incrementing the AFQT in predicting Soldier performance during training. On the retention-related criteria, the experimental predictors yielded incremental validity estimates typically in the .10s, and as high as .38 (an 800%+ gain in R over the AFQT). The percentage gains in R over the AFQT for predicting 6-month attrition were also significant. The experimental predictors incremented the AFQT by 66.7% (PSJT) to 285.5% (RBI) when predicting 6-month attrition. Across the retention-related criteria, the RBI generally emerged as the measure demonstrating the greatest gains over the AFQT, followed by the TAPAS, the AIM, and the WPA.

In regards to the subgroup differences analyses, the experimental predictors generally exhibited subgroup score differences (by gender, race, and ethnicity) that were about half the size, on average, of those observed on the AFQT. Further, on those measures or scales where there were sizeable subgroup differences, their direction was such that minority group members tended to score higher, on average, than majority group members. The exceptions to this finding were on scales measuring physically-oriented attributes, where one would reasonably expect to observe substantive gender differences on these attributes (e.g., the RBI's Fitness Motivation scale, the WPA Realistic Interest dimension scale, the WPA Mechanical and Physical facet scales).

Utilization and Dissemination of Findings:

These findings provide useful information to Army personnel managers and researchers about the potential of experimental predictor measures to increment the existing AFQT in selecting new Soldiers, in particular, measures assessing non-cognitive attributes. The Army Class longitudinal validation research will continue with the collection of in-unit job performance and retention data on participating Soldiers and implementation of additional selection criterion-related validation analyses as well as analyses to evaluate potential for MOS classification. The EEEM initiative will continue as a separate effort involving administration of selected experimental predictor measures to new Army applicants in an operational setting, as part of an Initial Operational Test and Evaluation (IOT&E) to start in May 2009.

VALIDATING FUTURE FORCE PERFORMANCE MEASURES (ARMY CLASS): END OF TRAINING LONGITUDINAL VALIDATION

CONTENTS

Page

CHAPTER 1: INTRODUCTION 1
Deirdre J. Knapp (HumRRO) and Tonia S. Heffner (ARI) 1
Background 1
Overview of the Army Class Research Program 2
Overview of Report 3

CHAPTER 2: LONGITUDINAL RESEARCH DESIGN 4
Deirdre J. Knapp (HumRRO) and Tonia S. Heffner (ARI) 4
Data Collection Points and Sample 4
Criterion Measures 4
Selection of Criterion Measures 4
Criterion Measure Development 5
Criterion Measure Descriptions 6
Predictor Measures 10
Selection of Predictor Measures 10
Description of Predictors 13

CHAPTER 3: DATA COLLECTION AND DATABASE DEVELOPMENT 16
Deirdre J. Knapp and Ani S. DiFazio (HumRRO) 16
Predictor Data Collections 16
Overview 16
Session Schedules 16
Training Criterion Data Collections 17
Overview 17
Session Schedules 18
Database Construction 18
Data Processing 19
Securing and Merging in Archival Data 19
Data Cleaning 19
Sample Descriptions 19
Predictor Sample 19
Training Criterion Sample 21

CONTENTS (continued)

Page

CHAPTER 4: MEASURE SCORING AND PSYCHOMETRIC PROPERTIES.....24
Matthew T. Allen, Yuqui A. Cheng, Michael J. Ingerick, and Joseph P. Caramagno (HumRRO)..... 24
Criterion Measure Scores and Associated Psychometric Properties 24
Job Knowledge Tests 24
Rating Scales..... 25
Army Life Questionnaire 25
Six-Month Attrition 26
IET School Performance and Completion 27
Predictor Measure Scores and Associated Psychometric Properties 28
Armed Services Vocational Aptitude Battery (ASVAB) 28
Assessment of Individual Motivation (AIM)..... 28
Tailored Adaptive Personality Assessment System (TAPAS-95s) 28
Rational Biodata Inventory (RBI)..... 28
Predictor Situational Judgment Test (PSJT)..... 29
Army Knowledge Assessment (AKA)..... 29
Work Preferences Assessment (WPA) 29

CHAPTER 5: ANALYSIS FINDINGS.....31
Michael J. Ingerick, Yuqui A. Cheng, and Matthew T. Allen (HumRRO)..... 31
Analysis Approach..... 31
Estimating the Incremental Validity of the Experimental Predictors 31
Estimating Subgroup Differences on the Experimental Predictors 32
Findings..... 32
Incremental Validity of the Experimental Predictor Measures..... 32
Subgroup Differences on the Experimental Predictors..... 40

CHAPTER 6: SUMMARY AND CONCLUSIONS 42
Michael J. Ingerick (HumRRO)..... 42
Summary of Main Findings 42
Incremental Validity 42
Subgroup Differences 42
Limitations and Issues..... 43
Comparing Results from the Army Class Longitudinal Validation to the Concurrent Validation..... 43
Generalizabilty of Findings to an Operational Setting 43
Future Research 44

CONTENTS (continued)

Page

REFERENCES.....45

APPENDIX A: DESCRIPTIVE STATISTICS AND SCORE INTERCORRELATIONS FOR SELECTED CRITERION MEASURES.....A-1

APPENDIX B: DESCRIPTIVE STATISTICS AND SCORE INTERCORRELATIONS FOR SELECTED PREDICTOR MEASURES.....B-1

APPENDIX C: SCALE-LEVEL CORRELATIONS BETWEEN SELECTED PREDICTOR AND CRITERION MEASURES.....C-1

APPENDIX D: PREDICTOR SCORE SUBGROUP DIFFERENCES.....D-1

List of Tables

Table 2.1. Summary of Longitudinal Validation Training Criterion Measures 5
Table 2.2. Description of the Army-Wide Performance Rating Scales (PRS) 7
Table 2.3. Description of the Training Army Life Questionnaire Scales 9
Table 2.4. Summary of Longitudinal Validation Predictor Measures 11
Table 2.5. Predictor Measures by Type and Characteristics Assessed 12

Table 3.1. Predictor Data Collection Session Schedules by Phase 17
Table 3.2. Schedule of Training Criterion Data Collection Sessions for Soldiers 18
Table 3.3. Predictor Sample by Phase and Reception Battalion 20
Table 3.4. Predictor Sample by MOS and Component 20
Table 3.5. Descriptive Statistics for Longitudinal Validation Predictor Sample 21
Table 3.6. Training Criterion Sample by MOS and Component 21
Table 3.7. Training Criterion Sample by MOS and Demographic Subgroup 22
Table 3.8. Archival Criterion Sample by MOS and Component 22
Table 3.9. Archival Criterion Sample by MOS and Demographic Subgroup 23

Table 4.1. Descriptive Statistics and Reliability Estimates for Job Knowledge Tests (JKTs) 24
Table 4.2. Attrition Rates through Six Months of Service by MOS 26
Table 4.3. Descriptive Statistics for Archival IET School Performance Criteria 27

CONTENTS (continued)

Page

Table 5.1. Incremental Validity Estimates and Predictive Validity Estimates for Experimental Predictors over the AFQT for Predicting Performance-Related Criteria (Continuous Criteria).....33
Table 5.2. Incremental Validity Estimates and Predictive Validity Estimates for Experimental Predictors over the AFQT for Predicting Disciplinary Incidents (Dichotomous).....35
Table 5.3. Incremental Validity Estimates and Predictive Validity Estimates for Experimental Predictors over the AFQT for Retention-Related Criteria (Continuous Criteria) 38
Table 5.4. Incremental Validity Estimates and Predictive Validity Estimates for Experimental Predictors over the AFQT for Predicting Retention-Based Criteria (Dichotomous Criteria).....39

Table A.1. Descriptive Statistics and Reliability Estimates for the Army-Wide (AW) and MOS-Specific Performance Rating Scales (PRS) 1
Table A.2. Intercorrelations among Army-Wide (AW) and MOS-Specific PRS..... 2
Table A.3. Descriptive Statistics and Reliability Estimates for the Army Life Questionnaire (ALQ) Scales by MOS..... 3
Table A.4. Intercorrelations among ALQ Scale Scores..... 5

Table B.1. Descriptive Statistics for the Armed Services Vocational Aptitude Battery (ASVAB) Subtests and Armed Forces Qualification Test (AFQT) 1
Table B.2. Intercorrelations among ASVAB Subtest and AFQT Scores 1
Table B.3. Descriptive Statistics and Reliability Estimates for Assessment of Individual Motivation (AIM) Scales..... 2
Table B.4. Intercorrelations among AIM Scales 2
Table B.5. Descriptive Statistics for Tailored Adaptive Personality Assessment System (TAPAS-95s) Scales..... 3
Table B.6. Intercorrelations among TAPAS-95s Scales..... 3
Table B.7. Descriptive Statistics and Reliability Estimates for Rational Biodata Inventory (RBI) Scale Scores..... 4
Table B.8. Intercorrelations among RBI Scale Scores 5
Table B.9. Descriptive Statistics and Reliability Estimates for Army Knowledge Assessment (AKA) Scales..... 6
Table B.10. Intercorrelations among AKA Scales..... 6
Table B.11. Descriptive Statistics and Reliability Estimates for Work Preferences Assessment (WPA) Dimension and Facet Scores 7
Table B.12. Intercorrelations among WPA Dimension and Facet Scores..... 8

CONTENTS (continued)

Page

Table C.1. Correlations between Predictor Scale Scores and Selected Performance-Related Criterion Measures....1
Table C.2. Correlations between Predictor Scale Scores and Selected Retention-Related Criterion Measures....4
Table C.3. Correlations between the AFQT and Scale Scores from the Experimental Predictor Measures....6
Table C.4. Correlations between Scales Scores from the TAPAS-95s and Other Temperament Predictor Measures....8
Table C.5. Correlations between Scale Scores from the WPA and the AKA....9
Table C.6. Correlations between Scale Scores from the TAPAS-95s and the WPA....10
Table C.7. Intercorrelations among Scale Scores from Selected Performance-Related Criterion Measures....11
Table C.8. Intercorrelations among Scale Scores from Selected Retention-Related Criterion Measures....11

Table D.1. Standardized Mean Differences (Cohen's d) by Subgroup Combination and Predictor Measure....1

List of Figures

Figure 2.1. Example Army-wide training rating scale.... 7
Figure 2.2. Example MOS-specific training criterion rating scale.... 8

VALIDATING FUTURE FORCE PERFORMANCE MEASURES (ARMY CLASS): END OF TRAINING LONGITUDINAL VALIDATION

CHAPTER 1: INTRODUCTION

Deirdre J. Knapp (HumRRO) and Tonia S. Heffner (ARI)

Background

The Personnel Assessment Research Unit (PARU) of the U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) is responsible for conducting manpower and personnel research for the Army. The focus of PARU's research is maximizing the potential of the individual Soldier through maximally effective selection, classification, and retention strategies, with an emphasis on the changing needs of the Army as it transforms into the future force.

The "Army Class" research program is a continuation of separate but related efforts that ARI has been pursuing since 2000 to ensure the Army is provided with the best personnel to meet the emerging demands of the 21st century. This research program is intended to support changes to the Army enlisted personnel selection and classification system that will result in improved performance, Soldier satisfaction, and service continuation. The current system relies primarily on the Armed Services Vocational Aptitude Battery (ASVAB), which is a cognitive aptitude test.

Army Class builds on three prior research efforts designed to improve the Army personnel system. These are *Maximizing Noncommissioned Officer (NCO) Performance for the 21st Century* (NCO21; Knapp, McCloy, & Heffner, 2004)*; New Predictors for Selecting and Assigning Future Force Soldiers* (Select21; Knapp, Sager, & Tremble, 2005); and *Performance Measures for 21st Century Soldier Assessment* (PerformM21; Knapp & Campbell, 2006). The NCO21 research was designed to identify and validate non-cognitive predictors of NCO performance for use in the junior NCO promotion system. The Select21 research was designed to provide new personnel tests to improve the ability to select and assign first-term Soldiers with the highest potential for future jobs. The Select21 effort validated new and adapted individual difference measures against criteria representing both "can do" and "will do" aspects of performance. The emphasis of the PerformM21 research project was to examine the feasibility of instituting routine competency assessments for enlisted personnel. As such, the researchers focused on developing cost-effective job knowledge assessments and examining the role of assessment within the overall structure of Army operational, education, and personnel systems. Because of their unique but complementary emphases, these three research efforts provide a strong theoretical and empirical foundation (including potential predictors and criteria) for the current project of examining enlisted personnel selection and classification.

The Army Class effort, formally titled *Validating Future Force Performance Measures*, began in 2006 with contract support from the Human Resources Research Organization (HumRRO). There is a 6-year plan for this research, as described next.

Overview of the Army Class Research Program

In the first year of the Army Class research program (2006), there were three distinct activities— one supporting military occupational specialty (MOS) reclassification of experienced Soldiers and two supporting pre-enlistment MOS classification. The idea behind the first activity was that job knowledge tests could potentially be used to facilitate reclassification of experienced Soldiers by assessing knowledge and skills applicable to their new MOS, then focusing retraining on areas of deficiency. The project team thus developed prototype job knowledge tests (JKTs) for several MOS (Moriarty, Campbell, Heffner, & Knapp, 2009). Given the resources required to conduct classification research in the Army that will support the needs of each of over 200 MOS, a second activity in Year 1 was to convene an expert panel to recommend strategies to make this goal more achievable for the Army (Campbell et al., 2007). Finally, the project team collected concurrent validation data using experimental pre-enlistment predictor measures and performance criterion measures developed and administered in the Select21 project (Knapp et al., 2005). The goal was to supplement the Select21 database to better support classification analyses. Although the results of these analyses were still based on generally small sample sizes and incumbent Soldiers, they indicated that the experimental predictor measures showed promise for enhancing the classification of entry-level Soldiers (Ingerick, Diaz, & Putka, 2009).

In Year 2 (2007), the emphasis of the Army Class research program was shifted to more fully focus on Soldier selection as well as classification issues. This emphasis was not only applied to the planned longitudinal criterion-related validation effort, which began in Year 2 with the administration of experimental predictor measures to over 11,000 new Soldiers, but was also reflected in the initiation of a companion ARI project entitled *Expanded Enlistment Eligibility Metrics (EEEM)*. The EEEM effort has a shorter timeframe for making recommendations to the Army about the use of new pre-enlistment tests to supplement the ASVAB. Additionally, the EEEM project led to the addition of two experimental pre-enlistment measures to the longitudinal research predictor set—an experimental version of the Assessment of Individual Motivation (AIM) and the Tailored Adaptive Personality Assessment System (TAPAS).

In Year 3 of the research program (2008), training performance criterion data were collected from the longitudinal validation sample. The database includes criterion measures adapted for this research as well as archival data on attrition and training course scores. For the Army Class longitudinal validation of selection measures, the analyses were geared to documenting the extent to which the experimental pre-enlistment measures from Select21 predicted training criteria using the full training criterion sample. For the EEEM portion of the research, the analyses were conducted earlier in the year using training criteria collected to that point. The goal was to identify predictors to recommend to the Army for use in an Initial Operational Test and Evaluation (IOT&E) starting early in 2009.

ARI plans for Year 4 (2009) include collection of job performance data from Soldiers in the longitudinal validation sample, most of who will have been working in their units for 14 to 18 months. The EEEM effort will diverge into support for the 3-year IOT&E. This will include programming the selected predictors into the computerized test platform used by the Military Entrance Processing Command (MEPCOM) and implementing an evaluation plan that includes

collecting training criterion data from Soldiers who are administered the predictors during pre-enlistment testing.

Years 5 and 6 (2010 and 2011) will include a second round of job performance data collection from Soldiers in the longitudinal validation sample. Most of the Soldiers will be approaching the end of their first term of enlistment so the data may help determine predictors for reenlistment. Year 6 also will include final documentation of the longitudinal validation and recommendations to be incorporated in the IOT&E

Overview of Report

The present report describes the Army Class longitudinal validation research design. It details the sample, data collection plan, and the selection and administration of predictor and training criterion measures. It describes database construction and the resulting analysis samples for the psychometric evaluation and training criterion-related validation analyses. A companion report (Knapp & Heffner, 2009) provides more detail on the EEEM portion of the research.

CHAPTER 2: LONGITUDINAL RESEARCH DESIGN

Deirdre J. Knapp (HumRRO) and Tonia S. Heffner (ARI)

This chapter describes the research design for the Army Class longitudinal validation, beginning with the sample selection strategy and plan for collecting data from participating Soldiers at up to four points in time. Selection, development, and descriptions of the training criterion measures and then the predictor measures are described.

Data Collection Points and Sample

In 2007 through early 2008, predictor data were collected from new Soldiers as they entered the Army through one of four Army reception battalions. Training performance criterion data were subsequently obtained on participating Soldiers at the completion of their Initial Entry Training (IET)—either Advanced Individual Training (AIT) or One-Station Unit Training (OSUT), as applicable to the MOS. This criterion data collection included only Soldiers who were in one of the six MOS-specific samples described below. The plan is to collect *job* performance criterion data from as many of the longitudinal validation Soldiers as possible at two points – in 2009 and again in 2010 when most Soldiers will have 2 to 3 years experience working in their units. This plan should thus yield data collected from at least a subset of the participating Soldiers at four different points in their Army careers.

Soldiers in the longitudinal predictor data collection were drawn from two types of samples: (a) MOS-specific samples targeting six entry-level jobs and (b) an Army-wide sample with no MOS-specific membership requirements. The six MOS-specific samples targeted the following occupations:

- 11B (Infantryman)
- 19K (Armor Crewman)
- 31B (Military Police)
- 63B (Light Wheel Vehicle Mechanic)
- 68W (Health Care Specialist)
- 88M (Motor Transport Operator)

These six target MOS, individually and collectively, were selected on the basis of multiple considerations, including but not limited to their importance to the Army's mission and priorities (e.g., as measured by the number of Soldiers in the MOS) and the feasibility of developing MOS-specific criterion measures for use in the research within the specified timeframe.

Soldiers in the longitudinal validation sample are inclusive of all Army components—Regular Army (RA), U.S. Army Reserve (USAR), and the U.S. Army National Guard (ARNG).

Criterion Measures

Selection of Criterion Measures

To obtain a comprehensive perspective on the extent to which Soldiers would be successful in the Army, the Army Class measures at all criterion points include job knowledge

tests (JKTs), supervisor performance ratings (plus peer ratings at the training criterion data collection point), and attitudinal data captured on a self-report questionnaire. The six JKTs used as training criteria were specifically written to reflect the knowledge and procedural content of the six target MOS (MOS-specific). The in-unit criterion data collection points will use a JKT that assesses general Soldiering knowledge and procedures (Army-wide) for all Soldiers as well as MOS-specific JKTs for Soldiers in the six target MOS. The rating scales for all three criterion data collection points include both Army-wide and MOS-specific dimensions (for Soldiers in the six target MOS). The attitudinal questionnaire is suitable for all Soldiers regardless of MOS. The end of training measures are supplemented with archival criterion indicators, most particularly continuation data, updated periodically throughout the course of the research.

Criterion Measure Development

Development and descriptive details for the in-unit performance criterion measures are discussed in Moriarty et al. (2009). Here we discuss the training criteria, which are summarized in Table 2.1.

Table 2.1. Summary of Longitudinal Validation Training Criterion Measures

Criterion Measure	Description
Computer-Administered	
MOS-Specific Job Knowledge Test (JKT)	Measures Soldiers' knowledge of the basic facts, principles, and procedures required of first-term Soldiers in a particular MOS (e.g., the major steps in loading a tank main gun, the main components of an engine). Each JKT consists of about 70 items representing a mix of item formats (e.g., multiple-choice, multiple-response, rank order, and drag and drop).
MOS-Specific and Army-Wide (AW) Performance Rating Scales (PRS)	Measures Soldiers' performance during AIT/OSUT on two categories of dimensions required of first-term Soldiers: (a) MOS-specific (e.g., performs preventive maintenance checks and services, troubleshoots vehicle and equipment problems) and (b) Army-wide (e.g., exhibits effort, supports peers, demonstrates physical fitness). The PRS were designed to be completed by the supervisors and peers of the Soldier being rated.
Army Life Questionnaire (ALQ)	Measures Soldiers' self-reported attitudes and experiences through the end of AIT/OSUT. The ALQ consists of 13 scales. The content of the 13 scales covers two general categories: (a) commitment and other retention-related attitudes towards the Army and MOS at the end of AIT/OSUT (e.g., perceived fit with Army; perceived fit with MOS) and (b) performance and adjustment during IET (e.g., adjustment to Army life, number of disciplinary incidents during IET).
Archival	
Attrition	Attrition data were obtained on participating Regular Army Soldiers through their first 6 months of service in the Army. These data were extracted from the Tier Two Attrition Screen (TTAS) database.
Initial Entry Training (IET) Performance and Completion	Operational IET performance and completion data were obtained from two Army administrative personnel databases: (a) Army Training Requirements and Resources System (ATRRS) and (b) Resident Individual Training Management System (RITMS). Soldier data on three IET-related criteria were extracted from these databases: (a) graduation from AIT/OSUT; (b) number of times recycled through AIT/OSUT; and (c) average AIT/OSUT exam grade.

We had limited time to prepare the training criterion measures since the original research plan did not include this data collection point and access to subject matter experts (SMEs) or Soldiers for development and pilot testing was also limited. Therefore, we constructed the training criterion measures by adapting measures that had been developed for Soldiers in units. These measures came from the Select21 and PerformM21 research previously cited, as well as the Army's Project A (Campbell & Knapp, 2001), a major selection and classification research project which was conducted in the 1980s and early 1990s. There was no opportunity to pilot test the training criterion measures, but each MOS proponent allowed us access to a cadre of five or so AIT/OSUT instructors to assist in measure development. We worked with these SMEs through a series of teleconferences supported by email exchanges of draft materials and information.

To create JKTs suitable for administration at the end of training, items developed for the in-unit criterion JKTs were reviewed with SMEs to purge content that is primarily learned on-the-job. Development of trainee rating scales started with the Select21 and Army Class concurrent validation scales (or Project A rating scales if the other were not available). We worked with SMEs to revise, delete, or add rating dimensions to make them suitable for trainees. Because we were planning to collect ratings from peers, it was also necessary to simplify the language and minimize the use of Army jargon. For the Army-wide performance ratings, we developed a set of rating dimensions and a bi-polar rating scale system with assistance from a panel of senior NCOs. We significantly simplified the rater training provided in previous data collections, making it short and focused. Finally, we developed a relatively short form of the Select21 Army Life Questionnaire tailored to the training environment. Development of the training criterion measures is described further in Moriarty et al. (2009).

Criterion Measure Descriptions

Job Knowledge Tests

Depending upon the MOS, the JKT items were drawn from items originally developed in PerformM21 (Knapp & Campbell, 2006), Select21 (Collins, Le, & Schantz, 2005), and Project A (Campbell & Knapp, 2001). Most of the training JKT items are in a multiple-choice format with two to four response options. However, other formats, such as multiple response (i.e., check all that apply), rank ordering, and matching are also used. The number of items on each of the six training JKTs range from 60 to 82. The items make liberal use of visual images to make them more realistic and to reduce reading requirements for the test.

Performance Rating Scales

The training-oriented Army-wide rating scales measure aspects of Soldier performance critical to all Soldiers, such as the amount of effort they exhibit, commitment to the Army, and personal discipline. These dimensions were identified by drawing from the content of (a) the IET critical incident dimensions from Select21 used to help develop the Predictor Situational Judgment Test (Knapp et al., 2005), (b) training rating dimensions from Project A (Campbell & Knapp, 2001), and (c) the basic combat training (BCT) rating scales developed by ARI (Hoffman, Muraca, Heffner, Hendricks, & Hunter, 2009). We used a relatively non-standard format for these scales. Seven of the eight dimensions had multiple rating scales, and there was a

single rating of "MOS Qualification and Skill" for a total of 21 individual ratings. Each response scale has a behavioral statement on the low end (rating of 1) and on the high end (rating of 5) as shown in Figure 2.1. The rating scale dimensions are described in Table 2.2.

C. Personal Discipline						
Behaves consistently with Army Core Values; demonstrates respect in word and actions towards superiors, instructors, and others; adheres to training behavior limitations (for example, use of cell phones and tobacco).						
Complains about requirements and directions; may delay or resist following directions.	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	Follows requirements and directions willingly.

Figure 2.1. Example Army-wide training rating scale.

Table 2.2. Description of the Army-Wide Performance Rating Scales (PRS)

Dimension	Description
Effort	Three-scale measure assessing Soldiers' persistence and initiative demonstrated when completing study, practice, preparation, and participation activities during AIT/OSUT (e.g., persisting with tasks, even when problems arose; paying attention in class and studying hard).
Physical Fitness and Bearing	Three-scale measure assessing Soldiers' physical fitness and effort exhibited to maintain self and appearance to standards (e.g., meeting or exceeding basic standards for physical fitness, dressing and carrying self according to standard).
Personal Discipline	Five-scale measure assessing Soldiers' willingness to follow directions and regulations and to behave in a manner consistent with the Army's Core Values (e.g., showing up on time for formations, classes, and assignments; showing proper respect for superiors).
Commitment and Adjustment to the Army	Two-scale measure assessing Soldiers' adjustment to the Army way of life and demonstrated progress towards the completion of the Soldierization process (e.g., taking on changes in plans or tasks with a positive attitude).
Support for Peers	Three-scale measure assessing Soldiers' support for and willingness to help their peers (e.g., offering assistance to peers that are ill, distressed, or failing behind; treating peers with respect, regardless of cultural, racial, or other differences).
Peer Leadership	Three-scale measure assessing Soldiers' proficiency in leading their peers when assigned to an AIT/OSUT leadership position, (e.g., gaining the cooperation of peers; taking on leader roles as assigned; giving clear directions to peers).
Common Warrior Tasks Knowledge and Skill	A single scale assessing Soldiers' proficiency in learning and demonstrating knowledge and skills in performing Common Tasks during Warrior Task/Drill training.
MOS Qualification Knowledge and Skill	A single scale assessing Soldiers' proficiency in learning and demonstrating the knowledge and skills required for MOS qualification during AIT/OSUT.

The format of the MOS-specific rating scales is different from that used in the Army-wide scales. Each rating scale measures a single aspect of MOS-specific performance and is rated on a 7-point response scale, as illustrated in Figure 2.2. The number of dimensions varies depending on the MOS, but ranges from five to eight. The dimensions and associated anchors were adapted from the most recent first-term Soldier performance rating scales available to the project team. In most cases, they came from the Select21 research (Keenan, Russell, Le, Katkowski, & Knapp, 2005).

A. Learns to Use Aiming Devices and Night Vision Devices

How well has the Soldier learned to engage targets with aiming devices, to zero sights, and to operate and maintain night vision devices?

1 2	3 4 5	6 7
– Is unable to engage targets with bore light and other aiming devices. – Cannot zero sights accurately, in daylight or at night; does not understand field zero.	– Is able to engage targets with bore light and other aiming devices with practice and coaching. – Zeroes sights accurately, but not quickly, both in daylight and at night; can apply field zero.	– Is extremely proficient in engaging targets with all types of aiming devices. – Zeroes sights quickly and accurately without assistance both in daylight and at night; applies field and expedient zero methods.

Figure 2.2. Example MOS-specific training criterion rating scale.

Army Life Questionnaire (ALQ)

The ALQ was designed to measure Soldiers' self-reported attitudes and experiences through the end of training. The original form of the ALQ was developed in the Select21 project (Van Iddekinge, Putka, & Sager, 2005). The end-of-training ALQ consists of 13 scales, summarized in Table 2.3. The content of the 13 scales falls into two general categories: (a) commitment and other retention-related attitudes towards the Army and MOS at the end of AIT/OSUT (e.g., perceived fit with Army; perceived fit with MOS) and (b) performance and adjustment during IET (e.g., adjustment to Army life, number of disciplinary incidents during IET). About half of the 58 items constituting the end-of-training ALQ were derived from earlier versions of the measure administered in Select21 and the Army Class concurrent validation. The other half consisted of new content that was developed for an AIT/OSUT setting.

Table 2.3. Description of the Training Army Life Questionnaire Scales

Scale	Description
Commitment and Retention-Related Attitudes	
Attrition Cognitions	Four-item scale measuring the degree to which Soldiers think about attriting before the end of their first-term (e.g., "How likely is it that you will complete your current term of service?").
Career Intentions	Five-item scale measuring Soldiers' intentions to re-enlist and to make the Army a career (e.g., "How likely is it that you will re-enlist in the Army?").
Army Fit	Six-item scale measuring Soldiers' perceived fit with the Army in general (e.g., "The Army is a good match for me.").
MOS Fit	Nine-item scale measuring Soldiers' perceived fit with their MOS (e.g., "My MOS provides the right amount of challenge for me.").
Normative Commitment	Five-item scale measuring Soldiers' feelings of obligation toward staying in the Army until the end of their current term of service (e.g., "I would feel guilty if I left the Army before the end of my current term of service.").
Affective Commitment	Seven-item scale measuring Soldiers' emotional attachment to the Army (e.g., "I feel like I am part of the Army 'family.'").
Initial Entry Training (IET) Performance and Adjustment	
Adjustment to Army Life	Nine-item scale measuring Soldiers' adjustment to life in the Army (e.g., "Looking back, I was not prepared for the challenges of training in the Army.").
Number of Disciplinary Incidents	Two-item measure (each item is segmented into multiple sub-questions) that asks Soldiers to self-report whether they had been involved in a series of disciplinary incidents (e.g., "While in the Army, have you ever been formally counseled for lack of effort?").
Last Army Physical Fitness Test (APFT) Score	Single-item asking Soldiers to self-report their most recent APFT score.
Number of IET Achievements	Two-item scale measuring the number of self-reported formal achievements a Soldier had earned during IET (e.g., "In AIT or OSUT, were you designated as part of the Fast Track Program?").
Number of IET Failures	Three-item scale measuring the number of self-reported repeats, recycles, or failures a Soldier had experienced during IET (e.g., "In BCT, OSUT, or AIT, did you ever have to retake the APFT to qualify for record?").
Self-Rated AIT/OSUT Performance	A set of scales asking Soldiers to rate their performance relative to the Soldiers they trained with along four dimensions – Physical Fitness, Discipline, Field Exercises, and Classroom and Instructional Modules – using a 4-point scale (1 = *Below Average [Bottom 30%]* to 4 = *Truly Exceptional [Top 5%]*).
Self-Ranked AIT/OSUT Performance	Single item asking Soldiers to rank-order their performance in AIT/OSUT on four dimensions – Physical Fitness, Discipline, Field Exercises, and Classroom and Instructional Modules – from the strongest (1) to the weakest (4).

Archival Criteria

Attrition

Attrition data were obtained on participating Soldiers through their first 6 months of service in the Army. The 6-month timeframe was selected because (a) it roughly corresponds to the completion of IET for most Soldiers in most MOS and (b) it balances the maturity of the attrition criterion (i.e., longer timeframes lead to more stable estimates) with the number of Soldiers on whom attrition data were available at the time the analyses were conducted. Attrition

information was extracted for participating Soldiers from the Two Tier Attrition Screen (TTAS) database maintained by the U.S. Army Accessions Command. For reasons explained later, the attrition analyses were limited to Regular Army Soldiers whose 6-month attrition status was known at the time the data were extracted.

IET Performance and Completion

IET performance and completion data were obtained from two administrative personnel databases: (a) Army Training Requirements and Resources Systems (ATRRS) and (b) Resident Individual Training Management System (RITMS). Soldier data on three IET-related criteria were constructed from data extracted from these databases: (a) graduation from AIT/OSUT, (b) number of times recycled through AIT/OSUT, and (c) average AIT/OSUT exam grade.

Predictor Measures

Selection of Predictor Measures

The Armed Forces Qualification Test (AFQT), an ASVAB composite score currently used as the primary cognitive screen for service in the U.S. military, served as the operational score against which the experimental predictors were evaluated.

Assembling Objects (AO) is now administered to U.S. military applicants as part of the ASVAB but until recently had not been used to screen or select applicants. Past research has shown that AO could supplement one or more of the existing ASVAB subtests in predicting entry-level Soldier performance, while potentially yielding lower gender differences than subtests measuring comparable abilities (Peterson et al., 1992; Russell, Reynolds, & Campbell, 1994). We included scores on the AO subtest as an experimental predictor to be evaluated in the Army Class research.[1]

The starting point for the identification and preparation of other experimental predictor measures for the longitudinal validation was the Army's Select21 project. Given the Army Class project's initial emphasis on classification, the original primary goal was to identify predictors likely to prove useful for classification purposes. The secondary goal was to assess selection-oriented predictors that needed additional research in a predictive validation (as opposed to concurrent validation) context.

We initially believed that identifying predictors for the longitudinal data collection would be a matter of balancing constraints on administration time, facilities, and equipment with the research priorities for individual instruments. Accordingly, we systematically characterized each instrument with regard to administration requirements (e.g., time, paper versus computer administration), predictive potential based on prior research, sensitivity to performance variation in concurrent versus predictive validation designs, and potential for response distortion in an operational setting. It soon became evident, however, that two logistical constraints—a 2-hour administration time limit and the requirement for paper-based administration (because of the large numbers of Soldiers to be tested in single sittings)—made selection of the predictors very

[1] AO is now included in the Two Tier Attrition Screen (TTAS) used to screen applicants who have not earned a high school degree.

simple. Several desirable predictor measures requiring computer administration (notably the Work Suitability Inventory [WSI], Work Values Inventory [WVI], and the Record of Pre-Enlistment Training and Experience [REPETE]) could not be included in the longitudinal administration plan, thus permitting all remaining measures to be selected.

After the Army Class predictor data collection was underway, the ARI EEEM project was initiated and resulted in the addition of two additional predictor measures—the AIM and TAPAS. As will be described in more detail in the next chapter, this was accomplished by temporarily suspending administration of some of the originally selected predictors while data from a sufficient number of new Solders were collected on the AIM and TAPAS.

Table 2.4 summarizes the predictor measures selected for inclusion in the joint Army Class/EEEM research. Table 2.5 provides a mapping of these predictor measures to characteristics identified as important to first-term Soldier performance and retention (Knapp & Tremble, 2007). The experimental measures cover all major knowledges, skills, and attributes (KSAs) of interest with the exception of work values. The Select21 measure designed to address this KSA, the WVI, could not be used because it must be administered by computer.

Table 2.4. Summary of Longitudinal Validation Predictor Measures

Predictor Measure	Description
Baseline Predictor	
Armed Forces Qualification Test (AFQT)	Measures general cognitive ability. The AFQT is a rationally weighted composite based on four Armed Services Vocational Aptitude Battery (ASVAB) subtests (Arithmetic Reasoning, Mathematics Knowledge, Word Knowledge, and Paragraph Comprehension). Applicants must meet a minimum score on the AFQT to enter the Army.
Cognitive Predictor	
Assembling Objects (AO)	Measures spatial ability. AO is currently administered as part of the ASVAB, but until recently had not been used to screen or select applicants. AO is now included in the Two Tier Attrition Screen (TTAS) used to screen applicants who have not earned a high school degree.
Temperament Predictors	
Assessment of Individual Motivation (AIM) – EEEM	Measures six temperament characteristics predictive of first-term Soldier attrition and performance (e.g., work orientation, dependability, adjustment). Each item consists of four behavioral statements. Respondents are asked to self-select which statement is most descriptive of them and which statement is least descriptive of them.
Tailored Adaptive Personality Assessment System (TAPAS-95s) – EEEM	Measures 12 dimensions or temperament characteristics predictive of first-term attrition and performance (e.g., dominance, attention-seeking, intellectual efficiency, physical conditioning). Uses a multidimensional pairwise preference (MDPP) format in which respondents indicate which of two statements is most like them.
Rational Biodata Inventory (RBI)	Measures 14 temperament and motivational characteristics important to entry-level Soldier performance and retention. Items ask respondents about their past behavior, experiences, and reactions to previous life events (e.g., the extent to which they enjoyed thinking about the "plusses and minuses" of alternative approaches to solving a problem).

Table 2.4. (Continued)

Predictor Situational Judgment Test (PSJT)	Measures respondents' judgment and decision-making proficiency across situations commonly encountered prior to or during the first enlistment term (e.g., dealing with a difficult co-worker). Each item consists of a description of a problem situation and a list of four alternative actions that the respondent might take in that situation. Respondents rate the effectiveness of each action.
Person-Environment (P-E) Fit Predictors	
Work Preferences Assessment (WPA)	Measures respondents' preferences for different kinds of work activities and settings offered by different jobs (e.g., working with others, repairing machines or equipment). Items ask respondents to rate how important a series of characteristics is to their ideal job. Content is based on Holland's (1997) theory of vocational personality and work environment.
Army Knowledge Assessment (AKA)	Measures respondents' understanding or expectations about the kinds of work activities and settings typically offered by the Army. Respondents are asked to read a brief description of six work settings and then rate the extent to which they think each setting describes the Army. Like the WPA, content is based on Holland's (1997) theory of vocational personality and work environment.

Table 2.5. Predictor Measures by Type and Characteristics Assessed

		Measure					
Attribute Type	Knowledge, Skill, or Attribute	ASVAB	PSJT	AIM	TAPAS	RBI	WPA/ AKA
Aptitude/ Declarative Knowledge	Reading Skill/ Comprehension	X					
	Basic Math Facility	X					
	General Cognitive	X					
	Spatial Relations	X					
	Basic Electronics Knowledge	X					
	Basic Mechanical Knowledge	X					
Procedural Knowledge & Skill	Self-Management Skill		X				
	Self-Directed Learning		X				
	Sound Judgment		X				
Temperament	Team Orientation		X				
	Agreeableness		X	X	X	X	
	Cultural Tolerance			X	X	X	
	Social Perceptiveness		X			X	
	Achievement Motivation		X	X	X	X	
	Self-Reliance				X		
	Affiliation				X		
	Potency			X	X	X	
	Dependability		X	X	X	X	
	Locus of Control				X	X	
	Intellectance			X	X		
	Emotional Stability			X	X	X	

Table 2.5. (Continued)

Interests	Realistic	X
	Investigative	X
	Artistic	X
	Social	X
	Enterprising	X
	Conventional	X
Values	Growth	
	Comfort	
	Stimulation	
	Status	
	Altruism	
	Self-Direction	

Description of Predictors

Armed Forces Qualification Test

The AFQT is a rationally weighted composite of four ASVAB tests (Arithmetic Reasoning, Math Knowledge, Word Knowledge, and Paragraph Comprehension). Scores on the AFQT reflect an applicant's standing on general cognitive ability and are one of the metrics, in addition to applicant's high school degree status, used to judge recruit potential. Examinees are classified into categories based on their AFQT percentile scores (Category I = 93-99, Category II = 65-92, Category IIIA = 50-54, Category IIIB = 31-49, Category IV = 10-30, Category V = 1-9). The AFQT served as the baseline against which the experimental predictors were to be evaluated.

Assembling Objects (AO)

AO is an ASVAB subtest that measures spatial ability and was first developed in Project A (Russell et al., 2001). The items are graphical in nature, requiring respondents to visualize how an object will look when its parts are put together correctly.

Assessment of Individual Motivation (AIM)

AIM was added to the Army Class longitudinal validation as part of the EEEM initiative. The original AIM was developed to address faking concerns with the otherwise promising Assessment of Background and Life Experiences (ABLE) developed in Project A (White & Young, 1998; White, Young, & Rumsey, 2001). The AIM uses a forced-choice format to reduce fakability and to improve the accuracy of the self-report information. Respondents are asked to self-select which statements are most and least descriptive of them. The AIM measures six temperament characteristics predictive of first-term Soldier attrition and performance: Dependability (Non-Delinquency), Adjustment, Physical Conditioning, Leadership, Work Orientation, and Agreeableness. Each item consists of four behavioral statements (i.e., tetrads). The AIM is currently used operationally by the Army to screen applicants who have not earned a high school degree. The version of AIM administered in this research has 30 items. Currently, the AIM is used operationally by the Army in the TTAS program to screen applicants who are not high school diploma graduates.

Tailored Adaptive Personality Assessment System (TAPAS-95s)

TAPAS-95s was also added to the Army Class project as part of the EEEM effort. It is a new 12 dimension, 95-item personality measure, developed by the Drasgow Consulting Group under the Army's Small Business Innovation Research (SBIR) program (Drasgow, Stark, & Chernyshenko, 2006; Stark, Chernyshenko, & Drasgow, 2008). The instrument builds on the foundational work of the AIM by incorporating features designed to promote resistance to faking and by including narrow personality constructs (i.e., facets) that are known or expected to predict outcomes in military settings. The TAPAS measures dimensions or temperament characteristics predictive of first-term Soldiers attrition and performance (e.g., Dominance, Attention-Seeking, Intellectual Efficiency, Physical Conditioning). The items are similar to those on the AIM, but use two statements instead of four. Respondents indicate which statement is most like them. The version of the TAPAS administered in the current research was a static, non-adaptive precursor to an item response theory (IRT)-based computerized adaptive personality assessment system capable of measuring up to 22 facets of the Big Five, as well as facets targeted to the military (e.g., physical conditioning).

Rational Biodata Inventory (RBI)

The RBI measures multiple temperament or motivational characteristics important to entry-level Soldier performance and retention (Kilcullen, Putka, McCloy, & Van Iddekinge, 2005). The measure has evolved in various ways depending on the application but grew out of the Assessment of Right Conduct (Kilcullen, White, Sanders, & Lazlett, 2003) and the Test of Adaptable Personality (Kilcullen, Mael, Goodwin, & Zazanis, 1999). Thus, with varying sets of items, it has been used in prior Army research and operational applications (e.g., for selection into Special Forces) for almost a decade. Items on the RBI ask respondents about their past behavior, experiences, and reaction to previous life events using Likert-style response options (e.g., the extent to which they enjoyed thinking about the plusses and minuses of alternative approaches to solving a problem). The RBI yields scores on a range of attributes (e.g., Achievement Motivation, Cognitive Flexibility, Fitness Motivation, Hostility to Authority, Peer Leadership, Self-Efficacy, and Stress Tolerance). The RBI used in the Army Class longitudinal validation has 101 items covering 14 attributes and is the same version used in the Select21 research (Kilcullen et al., 2005).

Predictor Situational Judgment Test (PSJT)

The PSJT is a 20-item paper-and-pencil measure designed to assess an individual's judgment and decision-making proficiency in challenging situations (e.g., working with uncooperative peers to accomplish a task; determining when to handle a problem alone versus consulting a supervisor; Waugh & Russell, 2005). The situations presented in the PSJT are civilian counterparts to the kinds of situations typically encountered by Soldiers during their first few months in the Army. These situations (and their underlying dimensions) were identified through collection of critical incidents from Soldiers in IET. Each item consists of a description of a situation followed by four actions that might be taken in that situation. Respondents rate the effectiveness of each action on a 1 to 7 scale (from "Ineffective" to "Very Effective"). The PSJT targets five kinds of situations or dimensions important to first-term Soldier performance: (a) Adaptability to Changing Conditions; (b) Relating to and Supporting Peers, (c) Teamwork, (d)

Self-Management, And (e) Self-Directed Learning. Although the PSJT items were written to reflect these dimensions, it is designed to yield a single total score.

Work Preferences Assessment (WPA)

The Work Preferences Assessment (WPA) is designed to assess an individual's preferences (or fit) for different kinds of work activities and environments (Van Iddekinge et al., 2005). The 72 items comprising the WPA were written to measure each of the six dimensions and their subfacets underlying Holland's (1997) theory of vocational personality and work environment. According to Holland's theory, work interests are expressions of personality that can be used to categorize individuals and work environments into six types (or dimensions): Realistic (R), Investigative (I), Artistic (A), Social (S), Enterprising (E), and Conventional (C). For each dimension or facet, the WPA contains three types of items: (a) interests in work activities (e.g., "A job that requires me to teach others"), (b) interests in work environments or settings (e.g., "A job that requires me to work outdoors"), and (c) interests in learning opportunities (e.g., "A job in which I can learn how to lead others"). Respondents are asked to rate each item in terms of its importance to their ideal job using a 5-point Likert-type scale (1 = "*Extremely unimportant to have in my ideal job*" to 5 = "*Extremely important to have in my ideal job*") (Putka & Van Iddekinge, 2007).

The WPA yields six dimension scores (corresponding to each of the six RIASEC dimensions) and 14 facet scores (corresponding to facets underlying the six RIASEC dimensions). These raw scores can then be combined or modified based on additional data to obtain multiple, alternative sets of scores for use in one or more of the Army's personnel management objectives.

Army Knowledge Assessment (AKA)

The Army Knowledge Assessment (AKA) is a 30-item instrument that assesses Soldiers' knowledge about the extent to which the current Army (in general) supports each RIASEC dimension (Van Iddekinge et al., 2005). Respondents are asked to read a brief description of six work settings and then rate the extent to which they think each setting describes the Army. The AKA yields six dimension scores, corresponding to the six RIASEC dimensions defined by Holland (1997). These raw scores can then be combined or modified based on additional data to obtain alternative sets of scores for use in one or more of the Army's personnel management objectives. Conceptually, the AKA is distinguished from the WPA in that it indicates whether respondents have realistic expectations about the interests that would be satisfied with Army life whereas the WPA indicates whether respondents are interested in what Army life offers. Both are strategies for predicting person-environment fit.

CHAPTER 3: DATA COLLECTION AND DATABASE DEVELOPMENT

Deirdre J. Knapp and Ani S. DiFazio (HumRRO)

In this chapter we describe both the predictor and training criterion data collections. We also describe the data processing, cleaning, and integration of archival data which determine the resulting Soldier samples.

Predictor Data Collections

Overview

Predictor data were collected from new Soldiers entering four reception battalions during the period of May 2007 through February 2008, ensuring that the resulting sample would reflect the recruit variations anticipated over the course of a year. Data collection visits were scheduled with each reception battalion to optimize our ability to gather data on Soldiers in the six target MOS as well as to maximize the total number of Soldiers tested. Data were collected over the course of 31 data collection site visits.

Data collections took place on weekends and were conducted by teams of ARI and HumRRO personnel. Sites were staffed with a minimum of two people. At Fort Jackson, where there were generally two rooms of Soldiers testing at once, we had teams of four to six people. We developed a Test Administrator (TA) Manual that was periodically updated to reflect changes in procedures (e.g., there was a TA manual specific to the October-November data collections in which the EEEM measures were administered). All data collectors participated in a training session prior to collecting data. The lead TA for each data collection prepared a test record to document the activities and issues related to each data collection which was used by those processing and scoring the data.

Session Schedules

At each reception battalion, data were collected from Soldiers in 2-hour test sessions. All sessions began with a project briefing and review of a Privacy Act statement. Soldiers then completed a Background Information Form that collected basic background information, such as MOS, race, ethnicity, and gender. After completing this form, Soldiers were administered the experimental predictor measures. Introduction of the EEEM measures into the data collection plan resulted in three phases of data collection in which the experimental predictor measures varied. Table 3.1 summarizes which experimental predictor measures were administered during each phase and the approximate time allotted for each measure. Note that three of the supporting reception battalions permitted 30 minutes of additional testing time during Phase 2, but one did not. At that location, we rotated skipping one measure in the instrument set to stay within the allotted 2-hour period.

Table 3.1. Predictor Data Collection Session Schedules by Phase

Phase 1 and 3 (May-September 2007; February 2008)		Phase 2 (October-November 2007)	
Activity	*Approximate Time Allotted*	*Activity*	*Approximate Time Allotted*
In-processing (seating, briefing, Privacy Act, BIF)	20 minutes	In-processing (seating, briefing, Privacy Act, BIF)	20 minutes
WPA	20 minutes	TAPAS-95s	30 minutes
AKA	10 minutes	WPA	20 minutes
PSJT	30 minutes	AKA	10 minutes
RBI	30 minutes	RBI	30 minutes
		AIM	30 minutes
Total	110 minutes		140 minutes

Note. Measures are presented in the order in which they were administered. BIF = Background Information Form, WPA = Work Preferences Assessment, AKA = Army Knowledge Assessment, PSJT = Predictor Situational Judgment Test, RBI = Rational Biodata Inventory, TAPAS-95s = Tailored Adaptive Personality Assessment System, AIM = Assessment of Individual Motivation.

Training Criterion Data Collections

Overview

We collected criterion data as Soldiers in the longitudinal validation target MOS completed AIT or OSUT. Thus, the training data collection schedule was driven by the flow of Soldiers in the predictor data collections and the length of training for each MOS. The data collections were conducted from mid-September 2007 through mid-August 2008.

To schedule suitable times to collect data at each training school, ARI provided the names of Soldiers we hoped to test and worked with the school point-of-contact to determine suitable dates for the 2-hour test sessions. We conducted 40 individual data collections across the six schools.

All sites except for Fort Benning were able to provide computer facilities. We stored a cache of ARI's laptop computers at Fort Benning to support the data collections for 11B Soldiers.

The training data collections were proctored by teams of two to three ARI and HumRRO staff members. As with the predictor data collections, we prepared a TA manual and provided training to data collection staff. Training included information about the instruments to be administered (including familiarization with the software for administering the measures), administration protocols, data documentation procedures, and materials/data handling procedures. During each data collection, the lead TA prepared the test session record.

Session Schedules

All Soldier sessions began with a project briefing and review of a Privacy Act statement. Soldiers then completed a computer-based Background Information Form. After finishing that form, Soldiers were administered computer-based versions of the MOS-specific JKT and ALQ, completing both measures at their own pace. Soldiers concluded the session by rating four to five peers on the AW and MOS-specific rating scales. Table 3.2 summarizes the schedule of the Soldier end-of-training criterion data collections.

Table 3.2. Schedule of Training Criterion Data Collection Sessions for Soldiers

Activity
In-Processing (seating, briefing, Privacy Act, Background Form)
MOS-Specific Job Knowledge Test (JKT)
Army Life Questionnaire (ALQ)
Peer Army-Wide (AW) and MOS-Specific Performance Rating Scales (PRS)

Note. Measures are listed in the order in which they were administered.

Soldiers' supervisors, typically the drill sergeant or an AIT/OSUT instructor, also provided performance ratings on Soldiers. Our goal was to collect two sets of supervisor ratings per Soldier. The supervisor sessions lasted one hour and took place concurrently with the Soldier data collections. During each session, supervisors were briefed on the project, reviewed a Privacy Act statement, completed a Background Information Form, and then rated upwards of 10 Soldiers on the Army-wide and MOS-specific rating scales. Prior to making their ratings, supervisors were asked to review the roster of participating Soldiers to indicate which Soldiers they would be able to rate. Supervisors completed the ratings on a computer and at their own pace. In some cases when the supervisors could not participate while the data collection staff was on site, we collected their ratings using a web-enabled application or on paper-based forms.

Database Construction

Constructing the predictor and training criterion validation database consisted of the following steps:

1. Processing the data.
2. Securing and merging in archival data from Army databases.
3. Cleaning the data.
4. Computing the scale scores and psychometric properties for the predictor and criterion measures.

Data Processing

In constructing the database to be used for all analyses, a number of steps were taken to ensure that the data were of the highest possible quality. Hard copy predictor data (particularly the background forms) were checked prior to electronic scanning to ensure that all Soldier responses were recorded by the scanner. For the training criterion data, initial processing involved uploading data to a central database and reading the data into an analyzable form. The population of Soldiers who completed each measure was electronically compared to the roster of Soldiers compiled in the field and inconsistencies in population membership were resolved. The logical consistency between records in a dataset and between variables within a dataset was investigated and corrections and edits were made as needed. Information from Test Session Logs and trip reports was culled to identify cases requiring a review and verification of their data.

Securing and Merging in Archival Data

Data collected in the field were merged with selected variables (e.g., ASVAB scores) extracted from Army databases, specifically the Enlisted Master File (EMF) and MEPCOM's Integrated Resource System (MIRS) on the predictor side, and the ATRRS, RITMS, and the TTAS attrition database on the criterion side. Data were retrieved from the Army databases by matching the Social Security Numbers (SSNs) of Soldiers participating in the data collections with SSNs in the Army databases.

Data Cleaning

After the data were processed and prepared by the database manager, the data were cleaned and screened to flag Soldiers with invalid or unusable data. Questionable Soldier responses (e.g., due to pattern responding) were dropped. The treatment of questionable data followed the same rules and protocols implemented in previous ARI research (e.g., Soldiers' data were excluded when they were missing more than 10% of the data for a scale or instrument) (Knapp & Tremble, 2007). Similar data checks and screens were applied to the archival data.

Sample Descriptions

Predictor Sample

Predictor data were collected on over 11,000 Soldiers. Descriptive information on the sample, post-data cleaning and scoring, is provided in Tables 3.3 through 3.5; the sample includes only those Soldiers who are non-prior service. Table 3.3 provides the number of Soldiers from whom predictor data were collected by phase and location. Table 3.4 describes the sample by MOS and component. Table 3.5 summarizes the demographic characteristics and entry qualifications of the sample.

Table 3.3. Predictor Sample by Phase and Reception Battalion

	Reception Battalion				
Phase	Fort Benning	Fort Jackson	Fort Knox	Fort Leonard Wood	Phase Totals
Phase 1 (May-Sep 2007)	618	1,865	380	885	3,748
Phase 2 (Oct-Nov 2007)	1,732	1,624	213	1,799	5,368
Phase 3 (Feb 2008)	451	442	438	367	1,698
Reception Battalion Totals	2,801	3,931	1,031	3,051	10,814

Note. The figures reported exclude Soldiers with prior military service.

Table 3.4. Predictor Sample by MOS and Component

	Component			
MOS	RA	ARNG	USAR	Totals
11B/X Infantryman	1,177	612	0	1,790
19K Armor Crewman	447	133	0	581
31B Military Police	616	580	288	1,484
63B Wheeled Vehicle Mechanic	186	181	105	472
68W Health Care Specialist	114	148	45	307
88M Motor Transport Operator	162	262	88	512
Army-Wide (AW)	2,668	1,873	1,113	5,654
Totals	5,370	3,789	1,639	10,800

Note. Fourteen Soldiers are missing MOS information and two are missing component information. The figures reported do not add up to the totals due to missing data. These data exclude Soldiers with prior military service.

As reported in Table 3.5, 19.5% of Soldiers participating in the predictor sample were female, with 14.1% identifying themselves as Black, and 7.6% as some other race. About 14% of the sample was Hispanic. In terms of quality as measured by AFQT Category, 32.2% of the Soldiers were Categories I-II, 24.7% were IIIA, 38.5% were IIIB, and 3.8% were IV. About 75% of Soldiers had earned a high school degree (or greater) at the time of accession. In general, these figures were comparable to those of Army Enlisted accessions, as a whole, based on FY 2006 numbers (Department of Defense, 2008) with few exceptions: (a) the predictor sample was somewhat more female; and (b) there were somewhat fewer AFQT Category I-II Soldiers and more IIIBs than in the full Army accession population.

Table 3.5. Descriptive Statistics for Longitudinal Validation Predictor Sample

	MOS							Totals	
Subgroup	11B/X	19K	31B	63B	68W	88M	AW	*n*	*%*
Gender									
Male	1,782	580	1,160	421	189	371	4,132	8,635	79.9
Female	0	0	321	47	117	140	1,486	2,111	19.5
Race									
White	1,558	499	1,281	387	251	374	4,070	8,420	77.9
Black	95	38	99	59	31	104	1,098	1,524	14.1
Other	129	42	99	24	25	32	467	818	7.6
Ethnicity									
White Non-Hispanic	1,383	465	1,176	351	231	352	3,572	7,530	69.6
Hispanic	272	59	202	57	41	48	847	1,526	14.1
AFQT Category									
I-II	539	168	446	106	232	131	1,855	3,477	32.2
IIIA	477	145	445	117	71	99	1,320	2,674	24.7
IIIB	710	244	576	213	3	225	2,188	4,159	38.5
IV	53	23	11	31	0	53	242	413	3.8
Highest Education Level (at Entry)									
HS Degree or Greater	1,183	376	1,230	310	256	388	4,349	8,092	74.8
No HS Degree	601	205	253	158	50	123	1,289	2,679	24.8
Totals	1,790	581	1,484	472	307	512	5,654	10,814	

Note. Fourteen Soldiers are missing MOS information. The figures reported do not add up to the totals due to missing data. These data exclude Soldiers with prior military service. Soldiers indicating more than one race are coded as "Other." The sample sizes for individual predictor measures vary due to missing data.

Training Criterion Sample

The training criteria were obtained from two primary sources. The first source was collected on-site from the Soldier and his or her supervisor(s) and peer(s). The second source was administrative records. The end of training criteria measures adapted for this research were administered to almost 2,400 Soldiers representing the Regular Army, Army Reserve, and Army National Guard, though approximately 100 of the Soldiers were not in our longitudinal sample. Tables 3.6 and 3.7 describe the criterion sample completing the training criteria measures following data cleaning and scoring; the sample includes only those Soldiers who were non-prior service and part of the predictor sample. Specifically, Table 3.6 describes the sample by MOS and component; Table 3.7 describes the demographics of the sample by MOS. Comparable information is provided for the archival criterion sample in Tables 3.8 and 3.9.

Table 3.6. Training Criterion Sample by MOS and Component

	Component			
MOS	RA	ARNG	USAR	Totals
11B	551	122	0	675
19K	354	113	0	470
31B	316	269	132	719
63B	102	78	40	222
68W	42	71	22	135
88M	23	35	15	73
Totals	1,388	688	209	2,294

Note. Nine Soldiers are missing component information. The figures reported do not add up to the totals due to missing data. These data exclude Soldiers with prior military service.

As shown in Table 3.7, 90.8% of the training criterion sample was male and 9.0% was female. About 86% of the sample was identified as White, 6.8% as Black, and 6.7% as some other race. In general, Soldiers in the sample were more likely to be male and less likely to be identified as a minority group member than the predictor sample. Same as the predictor sample, 14.1% of the training sample was Hispanic. The aptitude and educational qualifications of the training criterion sample were generally comparable, on average, to those in the predictor sample, with 72.7% of Soldiers having earned a high school degree or greater at the time of accession and the majority of Soldiers, 64.2%, being AFQT Category IIIA-IIIB (27.8% IIIA; 36.4% IIIB).

Table 3.7. Training Criterion Sample by MOS and Demographic Subgroup

	MOS						Totals	
Subgroup	11B	19K	31B	63B	68W	88M	*n*	*%*
Gender								
Male	674	470	583	204	91	61	2,083	90.8
Female	0	0	135	16	44	12	207	9.0
Race								
White	591	403	631	184	111	56	1,976	86.1
Black	36	30	41	28	12	10	157	6.8
Other	43	36	46	10	12	7	154	6.7
Ethnicity								
White Non-Hispanic	507	377	566	161	106	59	1,776	77.4
Hispanic	118	46	110	30	18	1	323	14.1
AFQT Category								
I-II	208	147	209	48	110	22	744	32.4
IIIA	197	125	224	52	24	15	637	27.8
IIIB	243	180	281	104	1	25	834	36.4
IV	24	18	4	15	0	11	72	3.1
Highest Education Level (at Entry)								
HS Degree or Greater	446	311	601	137	113	59	1,667	72.7
No HS Degree	228	160	118	83	22	14	625	27.2
Totals	675	470	719	222	135	73	2,294	

Note. The figures reported by subgroup and MOS do not add up to the totals due to missing data. These data exclude Soldiers with prior military service. Soldiers indicating more than one race are coded as "Other." The sample sizes for individual criterion measures vary due to missing data.

Table 3.8. Archival Criterion Sample by MOS and Component

	Component			
MOS	RA	ARNG	USAR	Totals
11B	944	479	0	1,424
19K	375	108	0	484
31B	558	521	277	1,356
63B	185	176	102	463
68W	114	145	45	304
88M	159	254	87	500
AW	2,609	1,792	1,080	5,481
Totals	4,944	3,475	1,591	10,012

Note. Fourteen Soldiers are missing MOS information and two Soldiers are missing component information. The figures reported do not add up to the totals due to missing data. These data exclude Soldiers with prior military service.

Table 3.9. Archival Criterion Sample by MOS and Demographic Subgroup

	MOS							Totals	
Subgroup	11B	19K	31B	63B	68W	88M	AW	*n*	*%*
Gender									
Male	1,423	484	1,063	416	188	364	3,994	7,932	79.1
Female	0	0	292	47	116	136	1,466	2,057	20.5
Race									
White	1,232	424	1,175	380	248	367	3,940	7,766	77.5
Black	75	27	92	57	31	101	1,068	1,451	14.5
Other	112	31	86	24	25	30	455	763	7.6
Ethnicity									
White Non-Hispanic	1,105	394	1,072	345	229	344	3,448	6,937	69.2
Hispanic	208	48	185	56	40	48	833	1,418	14.1
AFQT Category									
I-II	446	143	412	104	231	128	1,804	3,268	32.6
IIIA	383	117	411	116	70	95	1,284	2,476	24.7
IIIB	558	214	520	211	3	224	2,129	3,859	38.5
IV	35	10	9	31	0	50	231	366	3.7
Highest Education Level (at Entry)									
HS Degree or Greater	955	313	1,133	307	256	381	4,243	7,588	75.7
No HS Degree	469	171	223	156	48	119	1,236	2,422	24.2
Totals	1,424	484	1,356	463	304	500	5,481	10,026	

Note. The figures reported do not add up to the totals due to missing data. These data exclude Soldiers with prior military service. The sample sizes for individual criterion measures vary due to missing data.

CHAPTER 4: MEASURE SCORING AND PSYCHOMETRIC PROPERTIES

Matthew T. Allen, Yuqui A. Cheng, Michael J. Ingerick, and Joseph P. Caramagno (HumRRO)

In this chapter, we describe how the measures were scored and their psychometric properties as estimated in the Army Class sample. The criterion measures are presented first followed by the predictor measures.

Criterion Measure Scores and Associated Psychometric Properties

Job Knowledge Tests

A single, overall score was created for each JKT. Obtaining this score first involved computing and analyzing standard item statistics (e.g., *p*-values, item-total correlations) to identify poorly performing items. Poorly performing items were flagged and then reviewed by the lead JKT developer to make the final determination if the item should be dropped when computing a total score. Next, a raw total score was computed by summing the total number of points Soldiers earned across the final set of items retained for each JKT. All of the multiple-choice items were worth one point. Depending on the format of the non-traditional items (e.g., multiple response), they were worth one or more points. To facilitate comparisons across MOS, we computed a percent correct score based on the maximum number of points that could be obtained on each MOS test. For the criterion-related validity analyses, we converted the total raw score to a standardized score (or *z*-score) by standardizing the scores *within* each MOS.

Table 4.1 shows the descriptive statistics for the raw and percent correct scores, as well as internal consistency reliability estimates for the six MOS-specific JKTs. Based on percent correct scores, which ranged from 55.9% (63B) to 73.6% (68W), it is evident that the tests were fairly difficult, though not exceptionally so. The mean percent score across all six MOS tests was 62.15%. The internal consistency reliability estimates for the JKTs were acceptable, though the 19K estimate of .66 was a bit lower than would ordinarily be expected with this test method.

Table 4.1. Descriptive Statistics and Reliability Estimates for Job Knowledge Tests (JKTs)

MOS	*n*	Min	Max	Max Possible	*M*	*SD*	Mean Percent Correct	α
11B – Infantryman	629	42	91	118	69.91	9.51	59.1	.70
19K – Armor Crewman	432	18	54	60	38.01	6.09	63.4	.66
31B – Military Police	667	67	137	168	106.51	11.64	63.4	.72
63B – Light Wheel Vehicle Mechanic	202	31	99	122	68.09	12.56	55.9	.83
68W – Health Care Specialist	125	43	88	99	72.81	8.06	73.6	.73
88M – Motor Transport Operator	73	37	68	90	51.84	7.20	57.5	.77

Note. Max Possible = Maximum possible score on JKT; Percent Correct = Average percent correct received on JKT [*M* / Max Possible]; α = internal consistency reliability estimate (coefficient alpha).

Rating Scales

A single overall score was created for each Army-wide performance dimension and a composite of the MOS-specific scales. Computing these scores involved the following five steps. First, the ratings were screened to eliminate rater-ratee pairs with problematic data. This screening consisted of (a) checking the problem logs completed by the session proctors, (b) eliminating rater-ratee pairs where more than 10% of the ratings were missing, (c) eliminating rater-ratee pairs where the rater indicated "Not Applicable" on 50% or more of their ratings, and (d) eliminating rater-ratee pairs where the rater assigned the exact same profile of ratings to three or more of the Soldiers they rated.[2] Second, average peer rating scores on each scale were computed. For example, if a Soldier was rated by three peers, an average rating was created by computing a mean across the three raters. Third, average supervisor rating scores were computed using the same procedure as for the peer ratings. Fourth, peer and supervisor rating scale dimension scores were computed. This was done by taking the mean scores on all of the scales in a dimension (e.g., the three scales that describe *Effort* in the AW PRS), and computing an overall mean score. Finally, the peer and the supervisor ratings were again averaged to create a single overall rating for each dimension.

Descriptive statistics and estimates of interrater reliability for the AW PRS dimensions and MOS PRS composite scores are shown in Appendix A (Table A.1). The interrater reliability estimates were lower than desired, but consistent with our experience with the rating scales used in the Army Class and Select21 concurrent validations (Ingerick et al., 2009; Knapp & Tremble, 2007). Intercorrelations among the scales are provided in Table A.2. The 11B (Infantryman) MOS rating scale dimensions showed generally higher correlations with the Army-wide dimensions than the other MOS. The 68W (Medical Specialist) MOS ratings showed the lowest correlations with the Army-wide scales.

Army Life Questionnaire

Each ALQ scale was scored differently depending on the nature of the attribute being measured. For the self-evaluated IET performance scales, scores on the *Self-Rated AIT/OSUT Performance* and *Self-Ranked AIT/OSUT Performance* were left at the dimension-level and not aggregated to form a higher-order "self-rating" factor score. This was done because an examination of the intercorrelations suggested the scales were unique. Therefore, there were two scores (a ranking and a rating score) for four dimensions: Physical Fitness, Discipline, Field Exercises, Classroom and Instructional Modules. The *Last Army Physical Fitness Test (APFT) Score* was also unchanged. The *Number of Disciplinary Incidents, Number of IET Achievements,* and *Number of IET Failures* scales were scored by summing the number of "yes" responses to each item constituting the scale. The remaining seven scales - *Attrition Cognitions*, *Career Intentions*, *MOS Fit*, and *Army Fit*, *Normative Commitment*, *Affective Commitment*, and *Adjustment to Army Life* – were all scored with items that ranged from 1 (strongly disagree) to 5

[2] This last data screen only applied to Soldiers and Supervisors that had rated at least three Soldiers. Supervisors that rated more than 30 Soldiers were also exempted from this screen because they were likely to have assigned the same ratings to at least three Soldiers by virtue of the number of ratings that they completed. The data from Supervisors rating 30 or more Soldiers was examined closely, in combination with information from the problem logs and the other data screens, to ensure that their data were not problematic.

(strongly agree). Some of the items needed to be reverse-scored. Final scores were created for these remaining scales by computing the mean of the items.

Appendix A (Table A.3) shows descriptive statistics and internal consistency reliability estimates for the ALQ scores, by MOS and for the full sample. The reliability estimates were good (ranging from .79 to .94). Mean scores were generally similar across MOS. The Motor Transport Operators (88M) were on the higher end of the number of disciplinary incidents, but the mean number was still quite low. IET failures appeared to be most prevalent for Health Care Specialists (68W), as might be expected given the highly technical nature of this occupation. Score intercorrelations for the full sample are shown in Table A.4.

Six-Month Attrition

Only Soldiers that separated for applicable reasons were classified as attrits and included in our analyses. For the purposes of this research, attrition is a broad category that includes separations because of underage enlistment, conduct, family concerns, sexual orientation, drugs/alcohol, performance, physical standards/weight, mental disorder, or violations of the Uniformed Code of Military Justice. The reason for separation was determined by the Interservice Separation Code (ISC) associated with the Soldier. Once all of this information had been considered, a single 6-month attrition variable was computed. USAR and ARNG Soldiers were excluded from the attrition analysis because data on these Soldiers were incomplete and unreliable. Regular Army Soldiers whose attrition status was unknown at 6 months because they had insufficient time in service at the time the data extracted were also omitted from the analysis. Table 4.2 shows attrition rates for the total Regular Army sample and by MOS through 6 months of service, based on those Soldiers whose attrition status was known at the time the data were extracted.

Table 4.2. Attrition Rates through Six Months of Service by MOS

MOS	*N*	N_{Attrit}	$\%_{Attrit}$
Total Sample	4,478	539	12.0
MOS			
11B – Infantryman	931	190	20.4
19K - Armor Crewman	361	37	10.2
31B - Military Police	552	42	7.6
63B - Light Wheel Vehicle Mechanic	167	19	11.4
68W - Health Care Specialist	112	18	16.1
88M - Motor Transport Operator	140	19	13.6
AW - Army-Wide	2,215	214	9.7

Note. The statistics reported are based on Regular Army Soldiers only. *N* = number of Soldiers with 6-month attrition data at the time data were extracted. N_{Attrit} = number of Soldiers who attrited through 6 months of service. $\%_{Attrit}$ = percentage of Soldiers who attrited through 6 months of service [(N_{Attrit} /*N*) x 100].

IET School Performance and Completion

Data on IET school performance and completion were extracted from the ATTRS and RITMS databases. For the first variable, *Graduation from AIT/OSUT*, any Soldier who was discharged from Army during reception, basic training, or AIT/OSUT was coded as 0 (discharged). Any Soldier who graduated from AIT/OSUT was coded as 1 (graduated from AIT/OSUT). Any Soldier who was discharged during reception, basic training, or AIT/OSUT for nonpejorative, nonacademic reasons was coded as missing. The second variable, *Number of Recycles*, was created by counting total number of times a Soldier was recycled during IET. For the third variable*, Exam Grade*, the average score across all exam blocks during technical training was calculated for each Soldier and then standardized within an MOS.

Table 4.3 shows descriptive statistics for the graduation and recycle IET variables. The overall graduation rate was 88.4%, with the lowest rate being for 68W Soldiers (as also suggested by the related ALQ score). It is important to note that the IET data retrieved from archival sources was not complete. For example, although there were 10,814 Soldiers in the predictor sample, we retrieved graduation data on less than 7,000 and school exam scores on less than 1,500.

Table 4.3. Descriptive Statistics for Archival IET School Performance Criteria

Graduation from AIT/OSUT	*N*	N_{Grad}	$\%_{Grad}$
Total Sample	6,966	6,158	88.4
MOS			
11B – Infantryman	1,305	1,090	83.5
19K - Armor Crewman	413	413	100.0
31B - Military Police	1,306	1,236	94.6
63B - Light Wheel Vehicle Mechanic	328	295	89.9
68W - Health Care Specialist	40	18	45.0
88M - Motor Transport Operator	339	296	87.3
Number of Recycles through AIT/OSUT	*N*	*M*	*SD*
Total Sample	9,681	.09	.32
MOS			
11B – Infantryman	1,396	.10	.33
19K - Armor Crewman	475	.15	.38
31B - Military Police	1,333	.02	.14
63B - Light Wheel Vehicle Mechanic	459	.08	.29
68W - Health Care Specialist	297	.21	.48
88M - Motor Transport Operator	490	.10	.32

Note. *N* = number of Soldiers with data on the selected criterion. N_{Grad} = number of Soldiers who completed BCT and graduated from AIT/OSUT. $\%_{Grad}$ = percentage of Soldiers who completed BCT and graduated from AIT/OSUT [(N_{Grad} /*N*) x 100]. AIT = Advanced Individual Training; OSUT = One Station Unit Training.

Predictor Measure Scores and Associated Psychometric Properties

In this section, we describe how each of the Army Class predictor measures was scored and provide information about the psychometric properties of those scores.

Armed Services Vocational Aptitude Battery (ASVAB)

Soldiers' AFQT and ASVAB scores, including AO, were extracted from MEPCOM records and did not require any transformations or modifications. Descriptive statistics and score intercorrelations are provided in Appendix B (Tables B.1 and B.2, respectively).

Assessment of Individual Motivation (AIM)

For each AIM item tetrad, respondents provide two responses—one indicating the statement that is *most* like them and one indicating the statement that is *least* like them. A quasi-ipsative scoring method was used to generate four construct scores for each item (i.e. one score for each stem) based on whether the respondents indicate the stem is most like them, least like them, or neither. Scale scores were obtained by averaging–across items–the scores for stems measuring the same construct. A minimum of 80% of the items for any given construct must have been completed in order to obtain a score for that scale. Descriptive statistics and reliability estimates for the AIM scales are presented in Appendix B (Table B.3). The reliability estimates were all quite acceptable (ranging from .70 to .77). The validity (or lie scale) score was low, suggesting response distortion due to socially desirable responding was minimal.

Tailored Adaptive Personality Assessment System (TAPAS-95s)

For each TAPAS item pair, the respondents select the one item that is most like them. TAPAS-95s scoring was based on the multidimensional pairwise preference (MDPP) in which items were created by pairing statements subject to similarity constraints on social desirability and/or location (extremity). IRT was used to determine the dimension scores using the model originally proposed by Stark (2002). A detailed presentation of the scoring procedure is provided in the EEEM technical report (Knapp & Heffner, 2009). Descriptive statistics are shown in Appendix B (Table B.5) and scale intercorrrelations are shown in Table B.6.

Rational Biodata Inventory (RBI)

RBI scores were computed by summing responses to the items applicable to each scale (reverse-scoring as required) and dividing by the number of items in the scale. Substantive scale scores were not adjusted using the "Lie" scale score. Descriptive statistics and reliability estimates are shown in Appendix B (Table B.7). Most of the reliability estimates approached or exceeded .70. The substantive scales with fairly low internal consistency reliability estimates were Narcissism (.55) and Gratitude (. 43). These reliability estimates, as well as the mean scores, are generally similar to what was observed with the same version of the RBI used in the Select21 concurrent validation (Knapp & Tremble, 2007), with the highest score in both samples being Self-Efficacy and the lowest score being Hostility to Authority. Scale intercorrelations are provided in Table B.8.

Predictor Situational Judgment Test (PSJT)

For each PSJT item, the respondents rated the effectiveness of four possible actions in response to a hypothetical situation. The ratings were made on a 1 (ineffective) to 7 (very effective) response scale. The PSJT was scored in the manner developed and described by Waugh and Russell (2005). An initial judgment score for each response option was calculated using Equation 1 below.

$$Judgment\ Score_{Option\ x} = 6 - |\ Soldiers Rating_{Option\ x} - keyed Effectiveness_{Option\ x}\ | \qquad (1)$$

The keyed effectiveness ratings were based on judgments made by 67 SMEs during the Select21 project (Knapp & Tremble, 2007). We subtracted the difference between the respondent's rating and keyed effectiveness values from 6 to reflect the scores, so that higher values represented better scores. The judgment score for the entire test was the mean of the 80 option scores across the 20 scenarios. To minimize effects of a response pattern that recognizes that the keyed score will rarely be 1 or 7, the key was stretched as shown in Equations 2 and 3.

For original key values above 4.0, newValue = oldValue + 0.5 * (oldValue – 4). (2)
For original key values below 4.0, newValue = oldValue – 0.5 * (4 – oldValue). (3)

Finally, after stretching the key, we rounded the new value to the nearest integer. If the new value was less than one, we rounded it up to one; if the new value was greater than 7, we rounded it down to 7.

The mean PSJT score for the total sample was 4.67 (SD = .41, n = 4,970) and the coefficient alpha reliability estimate was .86. These results are very consistent with those obtained from the Army Class and Select21 concurrent validation samples (Ingerick et al., 2009; Waugh & Russell, 2005).

Army Knowledge Assessment (AKA)

The AKA yields six dimension scores corresponding to each of the six RIASEC dimensions. Items for each scale were averaged to create a total score for that scale. Total scores on each facet ranged from one to five. Descriptive statistics and reliability estimates for the AKA scales are shown in Table B.9. With the exception of Realistic Interests, which had a reliability estimate of .76, estimates for the remaining scales were high, ranging from .81 to .89. The scale with the highest mean score, not surprisingly for a sample of Soldiers, was Realistic Interests. AKA scale intercorrelations are shown in Table B.9

Work Preferences Assessment (WPA)

The WPA yields six raw dimension scores (corresponding to each of the six RIASEC dimensions) and 14 facet scores (corresponding to the subfacets underlying the six RIASEC dimensions). Raw scale scores were computed by obtaining the average of the scores across the items constituting each dimension or facet. Total raw scale scores range from one to five. Alternative algorithms for scoring the WPA are available, including algorithms that factor in environment or job-side data on the kinds of work activities and settings supported by the Army

in general or a specific job. Only the raw scale scores were used in the current research because (a) past research has shown that alternative scoring algorithms produce comparable criterion-related validity estimates and (b) the empirically-keyed scoring algorithms were developed under a concurrent validation design and using criterion data that were collected in-unit and not at the end of Soldiers' IET.

Descriptive statistics and reliability estimates for both the dimension and facet scores are shown in Table B.11 in Appendix B. Most reliability estimates are relatively high (mid-.70s to .90). Several of the facet scores were a bit lower, with Clear Procedures (a facet of Conventional Interests) being the score with the lowest estimated reliability (.64). The WPA score intercorrelations are shown in Table B.12.

CHAPTER 5: ANALYSIS FINDINGS

Michael J. Ingerick, Yuqui A. Cheng, and Matthew T. Allen (HumRRO)

This chapter describes the analyses examining the potential value of each of the experimental predictors to improve Army enlisted personnel selection decisions. In addition, we estimate the subgroup differences for each of the predictors as these differences also impact the potential operational value of each measure. Because the emphasis of the Army Class project at this stage was on selection, analyses examining the potential of the experimental predictors for improving MOS classification decisions were not conducted. Future plans call for conducting such analyses as MOS-specific sample sizes allow.

Analysis Approach

Estimating the Incremental Validity of the Experimental Predictors

To identify the measures that would best predict Soldier performance and retention, we estimated the incremental validity of the experimental predictor measures over the AFQT. Specifically, we fitted a series of hierarchical regression models, regressing each criterion measure onto Soldiers' AFQT scores in the first step, followed by the scale scores constituting a selected experimental predictor in the second step. Incremental validity is determined by estimating the increment in the multiple correlation (ΔR) when a new predictor is added to a baseline predictor(s) in a regression model. Consistent with the Army's personnel goals, we estimated the incremental validity of the experimental predictor measures over the AFQT for predicting both performance and retention-related criteria.

In estimating these models, we followed the same procedure used in analyzing data from previous research (e.g., the Select21, Army Class CV). This procedure was as follows:

1. Estimate the observed (uncorrected) multiple correlation *(R)* for the AFQT by regressing Soldiers' criterion scores on their AFQT scores.

2. Estimate *R* for AFQT and the new experimental predictor by regressing Soldiers' scores on the selected criterion onto AFQT scores and the scores for the new predictor measure (i.e., AFQT + Experimental Predictor).

3. Calculate the uncorrected incremental validity estimates (over AFQT) by subtracting the uncorrected (multiple) correlation obtained from Step 1 (the AFQT only) from the uncorrected multiple *R* (AFQT + New Experimental Predictor) obtained from Step 2.

Only the full scores for the experimental predictor measures were used when estimating these models. None of the experimental predictor scores used during estimation were optimally weighted or empirically keyed to a criterion.

To be consistent with the EEEM research, incremental validity was estimated using the observed (uncorrected) data. No corrections for statistical artifacts (criterion unreliability, range restriction, shrinkage or sample-specific error) were made when estimating incremental validity.

Estimating Subgroup Differences on the Experimental Predictors

Another important factor to be considered when evaluating the experimental predictor measures was subgroup differences. Subgroup differences represent the degree to which demographic subgroups score differently, on average, on a measure. Subgroup differences were examined by computing the standardized mean differences (i.e., Cohen's *d*) between targeted demographic subgroups on the scale scores constituting the experimental predictor measures. The demographic subgroups targeted for our analyses were (a) gender (female versus male), (b) race (Black versus White), and (c) ethnicity (Hispanic versus White, Non-Hispanic). Standardized mean differences were computed using a variant of Cohen's *d* statistic, where

$$d = (M_{COMPARISON} - M_{REFERENT})/SD_{REFERENT}.$$ [3]

For the purpose of this analysis, the referent group is the group that does not have special protections under relevant employment laws (e.g., males and Whites). Accordingly, the referent groups were Males, Whites, and Non-Hispanics; while the comparison groups were Female, Black, and Hispanic. All standardized mean differences were computed using the observed (uncorrected) data. No corrections for statistical artifacts were made when computing these differences.

Findings

Incremental Validity of the Experimental Predictor Measures

Tables 5.1 through 5.4 show uncorrected incremental validity estimates for the experimental predictor measures by criterion type for the full sample.[4] Based on theory and recent research examining the experimental predictor measures (e.g., Campbell, McCloy, Sager, & Oppler, 1993; Ingerick et al., 2009; Knapp & Tremble, 2007), we expected the incremental validity of the experimental predictor measures to vary by criterion type. For this reason, we first present the results based on the performance-related criteria, followed by those based on the retention-related criteria.

Predicting Technical and Non-Technical Performance-Related Criteria

Table 5.1 reports the incremental validity estimates for the experimental predictor measures over the AFQT for predicting continuously-scaled performance-related criteria, while Table 5.2 shows the incremental validity estimates over the AFQT for predicting a dichotomously-scaled criterion. Examination of Tables 5.1 and 5.2 reveals the following:

The experimental predictor measures demonstrated the potential to increment the AFQT in predicting a job knowledge-based performance criterion. The predictive validity of the AFQT for predicting MOS-specific JKT performance was high (R = .44). Nevertheless, the addition of

[3] M = Group Mean, SD = Group Standard Deviation.
[4] See Appendix C for the uncorrected scale-level correlations between selected predictor and criterion measures.

the experimental measures evidenced potentially non-trivial increments in prediction (Δ*Rs* = .01 to .04). Among the experimental predictors, the RBI exhibited the greatest gain over the AFQT (Δ*R* = .04), followed by the WPA facets (Δ*R* = .03), the TAPAS (Δ*R* = .03), the WPA dimensions and the PSJT (Δ*R* = .02), AO (Δ*R* = .02), and the AKA and AIM (Δ*R* = .01).

Table 5.1. Incremental Validity Estimates and Predictive Validity Estimates for Experimental Predictors over the AFQT for Predicting Performance-Related Criteria (Continuous Criteria)

Criterion/Predictor	*n*	AFQT Only	AFQT + Predictor	Δ*R*
MOS-Specific Job Knowledge Test (JKT)				
AO [1]	1,908	**.436**	**.453**	**.017**
AIM [6]	636	**.436**	**.448**	.012
TAPAS [12]	781	**.436**	**.461**	**.025**
PSJT [1]	1,308	**.436**	**.456**	**.020**
RBI [14]	1,639	**.436**	**.475**	**.039**
AKA [6]	2,001	**.436**	**.448**	**.012**
WPA Dimensions [6]	1,977	**.436**	**.460**	**.024**
WPA Facets [14]	1,976	**.436**	**.470**	**.034**
MOS-Specific Performance Ratings Composite				
AO [1]	2,042	**.148**	**.172**	**.024**
AIM [6]	676	**.148**	**.208**	**.060**
TAPAS [12]	837	**.148**	**.193**	.045
PSJT [1]	1,390	**.148**	.154	.006
RBI [14]	1,725	**.148**	**.194**	**.046**
AKA [6]	2,125	**.148**	.157	.009
WPA Dimensions [6]	2,098	**.148**	.157	.009
WPA Facets [14]	2,097	**.148**	.172	.024
Effort Ratings Composite (Army-Wide)				
AO [1]	2,080	**.189**	**.221**	**.032**
AIM [6]	687	**.189**	**.271**	**.082**
TAPAS [12]	846	**.189**	**.255**	**.067**
PSJT [1]	1,423	**.189**	**.203**	**.014**
RBI [14]	1,764	**.189**	**.239**	**.051**
AKA [6]	2,170	**.189**	.201	.012
WPA Dimensions [6]	2,142	**.189**	.199	.011
WPA Facets [14]	2,141	**.189**	**.217**	**.029**
Physical Fitness and Bearing Ratings Composite (Army-Wide)				
AO [1]	2,080	**.089**	**.141**	**.052**
AIM [6]	687	**.089**	**.295**	**.206**
TAPAS [12]	846	**.089**	**.258**	**.169**
PSJT [1]	1,423	**.089**	.090	.001
RBI [14]	1,764	**.089**	**.293**	**.204**
AKA [6]	2,170	**.089**	.106	.017
WPA Dimensions [6]	2,142	**.089**	**.118**	**.029**
WPA Facets [14]	2,141	**.089**	**.141**	**.052**

Table 5.1. (Continued)

Criterion/Predictor	*n*	AFQT Only	AFQT + Predictor	Δ*R*
Support for Peers Ratings Composite (Army-Wide)				
AO [1]	2,080	**.155**	**.179**	**.024**
AIM [6]	687	**.155**	**.256**	**.101**
TAPAS [12]	846	**.155**	**.239**	**.084**
PSJT [1]	1,423	**.155**	**.164**	**.009**
RBI [14]	1,764	**.155**	**.205**	**.051**
AKA [6]	2,170	**.155**	.167	.012
WPA Dimensions [6]	2,142	**.155**	**.180**	**.025**
WPA Facets [14]	2,141	**.155**	**.191**	**.036**
Peer Leadership Ratings Composite (Army-Wide)				
AO [1]	2,077	**.151**	**.180**	**.029**
AIM [6]	687	**.151**	**.266**	**.115**
TAPAS [12]	845	**.151**	**.230**	**.079**
PSJT [1]	1,421	**.151**	.159	.008
RBI [14]	1,762	**.151**	**.240**	**.089**
AKA [6]	2,167	**.151**	.160	.009
WPA Dimensions [6]	2,140	**.151**	.160	.009
WPA Facets [14]	2,139	**.151**	**.181**	.030
Personal Discipline Ratings Composite (Army-Wide)				
AO [1]	2,077	**.185**	**.206**	**.021**
AIM [6]	687	**.185**	**.298**	**.113**
TAPAS [12]	846	**.185**	**.310**	**.125**
PSJT [1]	1,423	**.185**	**.192**	**.007**
RBI [14]	1,764	**.185**	**.249**	**.064**
AKA [6]	2,170	**.185**	.190	.005
WPA Dimensions [6]	2,142	**.185**	**.205**	**.020**
WPA Facets [14]	2,141	**.185**	**.224**	**.039**
Army Physical Fitness Test (APFT) Score (Self-Reported)				
AO [1]	2,010	**.054**	.056	.002
AIM [6]	660	.054	**.350**	**.296**
TAPAS [12]	824	.054	**.336**	**.282**
PSJT [1]	1,391	**.054**	.057	.003
RBI [14]	1,722	**.054**	**.404**	**.350**
AKA [6]	2,106	**.054**	.073	.019
WPA Dimensions [6]	2,188	**.054**	**.100**	**.046**
WPA Facets [14]	2,187	**.054**	**.162**	**.108**

Note. AFQT = Armed Forces Qualification Test. AFQT Only = Correlation between the AFQT and the criterion. AFQT + Predictor = Multiple correlation (*R*) between the AFQT and the selected predictor measure with the criterion. Δ*R* = Increment in *R* over the AFQT from adding the selected predictor measure to the regression model (AFQT + Predictor – AFQT Only). Estimates in bold were statistically significant, $p < .05$ (two-tailed). The numbers in brackets after the title of the predictor measure indicate the number of scale scores that the measure contributed to the regression model. The WPA yields six dimension and 14 facet scale scores. Pairwise deletion was used to account for missing data.

Table 5.2. Incremental Validity Estimates and Predictive Validity Estimates for Experimental Predictors over the AFQT for Predicting Disciplinary Incidents (Dichotomous)

	n	AFQT Only	AFQT + Predictor	Δ*R*
Predictor				
AO [1]	2,019	**.103**	**.117**	**.015**
AIM [6]	659	**.130**	.204	.074
TAPAS [12]	824	.078	.188	.110
PSJT [1]	1,394	**.096**	**.116**	**.020**
RBI [14]	1,726	**.114**	**.224**	**.110**
AKA [6]	2,108	**.104**	.118	.013
WPA Dimensions [6]	2,090	**.100**	.115	.015
WPA Facets [14]	2,089	**.101**	.145	.044

Note. The effect sizes reflect Nagelkerke's *R*. Estimates in bold were statistically significant, $p < .05$ (two-tailed). The numbers in brackets after the title of the predictor measure indicate the number of scale scores that the measure contributed to the regression model. The WPA yields six dimension and 14 facet scores. Listwise deletion was used to account for missing data.

Selected experimental predictor measures exhibited significant potential to increment the prediction of ratings of MOS-specific performance over the AFQT. The predictive validity of the AFQT was lower for ratings of MOS-specific performance ($R = .15$) than for the MOS-specific JKT. Among all the experimental predictors, only the AIM ($\Delta R = .06$), the RBI ($\Delta R = .05$), and AO ($\Delta R = .02$), exhibited statistically significant incremental validity over the AFQT. However, since the predictive validity of the AFQT was relatively low, adding experimental predictors to the models resulted in a relatively large gain in *R*, even though the results were not technically statistically significant (e.g., TAPAS). For example, the TAPAS showed a 30.5% gain over the AFQT ($\Delta R = .05$). In general, the predictive validity estimates of the AFQT and the experimental predictors tended to be lower on the performance ratings than on the other criteria. This can be attributed to the low levels of interrater reliability observed for these measures.[5] However, this should not adversely influence the conclusions made about the relative predictive validity of the different measures.

With the exception of the AKA and WPA (dimensions), the experimental predictor measures emerged as useful predictors of Soldiers' effort ratings. The predictive validity of the AFQT was moderate ($R = .19$). Among the experimental predictors, the AIM ($\Delta R = .08$), TAPAS ($\Delta R = .06$), and RBI ($\Delta R = .05$) showed highest levels of incremental validity. These three measures exhibited a 43.5%, 35.5%, and 26.8% gain over the AFQT, respectively. The AO ($\Delta R = .03$), WPA (facets) ($\Delta R = .03$), and PSJT ($\Delta R = .01$) demonstrated relatively lower but still statistically significant incremental validity. These three measures exhibited a 17.1%, 15.3%, and 7.7% gain over the AFQT, respectively.

[5] The single-rater reliability estimates [ICC(A,1)] on the performance ratings ranged from .17 to .31, while the multi-rater reliability estimates [ICC(A,k)] ranged from .37 to .63, among the total sample. See Appendix A for a reporting of the interrater reliability estimates by scale and MOS. These estimates were comparable to those obtained in previous Army research (cf. Ingerick et al., 2008; Knapp & Tremble, 2007) and in the applied organizational research on performance ratings in general (cf. Visweveran, Ones, & Schmidt, 1996).

With the exception of the AKA and PSJT, the experimental predictor measures generally exhibited statistically significant incremental validity over the AFQT in predicting ratings of Soldiers' physical fitness and bearing. The predictive validity of the AFQT for Soldiers' physical fitness and bearing ratings was statistically significant, but relatively low in magnitude (R = .09). The AIM (ΔR = .21), RBI (ΔR = .20), and TAPAS (ΔR = .17) showed substantial levels of incremental validity, resulting in a 232%, 229%, and 190% gain over the AFQT, respectively. The AO (ΔR = .05), WPA (facets) (ΔR = .05), and WPA (dimension) (ΔR = .03) showed significant but lower levels of incremental validity.

The experimental predictor measures generally exhibited statistically significant incremental validity over the AFQT in predicting ratings of Soldiers' support for peers. The AFQT showed a moderate level of predictive validity (R = .16). With the exception of the AKA, all experimental predictors explained additional variance in the criterion that was statistically significant. The AIM (ΔR = .10), TAPAS (ΔR = .08), and RBI (ΔR = .05) evidenced the greatest incremental validity, and led to a 65.6%, 54.4%, and 32.9% gain over the AFQT, respectively. The WPA (dimensions and facets), AO, and PSJT demonstrated significant but lower level of incremental validity (ΔR = .01 to .04).

The experimental predictor measures exhibited some potential to increment prediction in ratings of peer leadership over the AFQT. The AFQT showed a moderate level of predictive validity (R = .15). Three experimental predictors, the AIM (ΔR = .12), RBI (ΔR = .09), and TAPAS (ΔR = .08) showed highest levels of incremental validity. The gain over the AFQT for these three predictors was 76.3%, 59.2%, and 52.1%, respectively. The AO (ΔR = .03) and WPA facets (ΔR = .03) evidenced lower but still significant increment in leadership ratings over AFQT. All other predictors (PSJT, AKA, WPA dimensions) added very little to the prediction of leadership ratings (ΔR < .01).

With the exception of the AKA, the experimental predictor measures exhibited statistically significant levels of incremental validity over the AFQT in predicting Soldiers' ratings of discipline. The AFQT showed a moderate level of predictive validity (R = .19). Among experimental predictors, the TAPAS (ΔR = .13), AIM (ΔR = .11), and RBI (ΔR = .06) evidenced the greatest incremental validity and demonstrated 67.8%, 61.0%, and 34.6% gains over the AFQT respectively. The AO (ΔR = .02), WPA facets, WPA dimensions, and PSJT only showed statistically significant but lower levels of incremental validity over the AFQT (ΔR = .01 to .04).

With the exception of the AO, AKA and PSJT, the experimental predictor measures exhibited statistically significant incremental validity over the AFQT in predicting Soldiers' Army Physical Fitness Test (APFT) score. The AFQT showed a limited level of predictive validity (R = .05) in predicting APFT score. On the contrary, some experimental predictors showed high levels of predictive validity. The RBI demonstrated the greatest incremental validity (ΔR = .35), followed by the AIM (ΔR = .30) and TAPAS (ΔR = .28). The WPA facets (ΔR = .11) and WPA dimensions (ΔR = .05) showed smaller, but still substantial, incremental validity.

Finally, as the only dichotomously scaled performance-related criterion, disciplinary incidents were significantly predicted by three experimental predictors: RBI, PSJT, and AO. The AFQT evidenced limited potential in predicting disciplinary incidents (R = .08 to .13). Among all the experimental predictors, the RBI (ΔR = .11) showed highest level of incremental validity. The PSJT (ΔR = .02) and AO (ΔR = .02) demonstrated low but statistically significant incremental validity coefficients. The gain in incremental validity over AFQT was 96.0%, 21.0%, and 14.2% for the RBI, PSJT and AO respectively. Other predictors, although not statistically significant, demonstrated substantial gains over AFQT. For example, 141% for TAPAS, 56.8% for AIM, 43.6% for WPA facets.

In sum, with the exceptions of the PSJT and the AKA, the experimental predictors demonstrated significant incremental validity over the AFQT for predicting performance-related criteria. The RBI demonstrated significant incremental validity over AFQT for all nine performance-related criteria. The AIM and AO subtest demonstrated incremental validity for eight out of nine performance-related criteria. The TAPAS exhibited incremental validity for seven out of nine criteria, the WPA facets six out of nine, and the WPA dimensions and PSJT five out of nine. AKA appeared to be the weakest predictor because it demonstrated incremental validity over AFQT for only one criterion (MOS-specific JKT).

Predicting Retention-Related Criteria

Table 5.3 reports the incremental validity estimates for the experimental predictor measures over the AFQT in predicting continuously-scaled retention-related criteria. Table 5.4 shows the incremental validity estimates of the experimental predictors over the AFQT in predicting 6-month attrition. Examination of Tables 5.3 and 5.4 reveals the following:

With the exception of AO and PSJT, the experimental predictor measures generally exhibited substantial potential to increment prediction over the AFQT for affective commitment. The predictive validity of the AFQT for predicting affective commitment was fairly low (R = .06). Among all of the experimental predictors, the RBI (ΔR = .35) evidenced the highest incremental validity over the AFQT in predicting affective commitment, followed by the WPA facets (ΔR = .19), TAPAS (ΔR = .19), AO (ΔR = .18), WPA dimensions (ΔR = .18), and AKA (ΔR = .14). The percentages of gain over AFQT were also substantial: 528% for RBI, 291% for WPA facets, 280% for TAPAS, 274% for AIM, 265% for WPA dimensions, and 212% for AKA.

With the exception of AO, the experimental predictor measures exhibited significant potential to increment over the AFQT for needs-supplies Army fit. The AFQT did not predict variance in needs-supplies Army fit (R = .01), but experimental predictors showed strong prediction which resulted in high incremental validity. The RBI (ΔR = .38) again evidenced the highest incremental validity. Other experimental predictors such as the TAPAS (ΔR = .26), AIM (ΔR = .26), WPA facets (ΔR = .25), WPA dimensions (ΔR = .22), and AKA (ΔR = .18) also demonstrated substantial incremental validity. Compared with other experimental predictors, the PSJT was a relatively weak predictor, but still reached the threshold of statistical significance (ΔR = .05).

Table 5.3. Incremental Validity Estimates and Predictive Validity Estimates for Experimental Predictors over the AFQT for Retention-Related Criteria (Continuous Criteria)

Criterion/Predictor	*n*	AFQT Only	AFQT + Predictor	Δ*R*
Affective Commitment				
AO [1]	2,001	**.066**	.067	.001
AIM [6]	659	.066	**.247**	**.181**
TAPAS [12]	824	.066	**.251**	**.185**
PSJT [1]	1,380	**.066**	.074	.008
RBI [14]	1,714	**.066**	**.415**	**.349**
AKA [6]	2,098	**.066**	**.206**	**.140**
WPA Dimensions [6]	2,077	**.066**	**.241**	**.175**
WPA Facets [14]	2,076	**.066**	**.258**	**.192**
Needs-Supplies Army Fit				
AO [1]	2,000	.012	.027	.016
AIM [6]	653	.012	**.269**	**.258**
TAPAS [12]	818	.012	**.273**	**.261**
PSJT [1]	1,387	.012	**.063**	**.051**
RBI [14]	1,718	.012	**.387**	**.375**
AKA [6]	2,096	.012	**.187**	**.175**
WPA Dimensions [6]	2,079	.012	**.233**	**.221**
WPA Facets [14]	2,078	.012	**.262**	**.251**
Career Intentions				
AO [1]	1,998	**.047**	.057	.011
AIM [6]	654	.047	**.259**	**.212**
TAPAS [12]	818	.047	**.298**	**.251**
PSJT [1]	1,387	.047	.053	.006
RBI [14]	1,711	.047	**.318**	**.271**
AKA [6]	2,093	**.047**	**.160**	**.113**
WPA Dimensions [6]	2,073	**.047**	**.192**	**.146**
WPA Facets [14]	2,072	**.047**	**.225**	**.178**
Attrition Cognitions				
AO [1]	1,997	**.048**	.049	.001
AIM [6]	657	.048	**.208**	**.160**
TAPAS [12]	815	.048	**.234**	**.186**
PSJT [1]	1,382	.048	.065	.018
RBI [14]	1,714	**.048**	**.304**	**.256**
AKA [6]	2,091	**.048**	**.158**	**.110**
WPA Dimensions [6]	2,072	**.048**	**.165**	**.118**
WPA Facets [14]	2,071	**.048**	**.207**	**.159**

Note. AFQT = Armed Forces Qualification Test. *AFQT Only* = Correlation between the AFQT and the criterion. *AFQT + Predictor* = Multiple correlation (*R*) between the AFQT and the selected predictor measure with the criterion. Δ*R* = Increment in *R* over the AFQT from adding the selected predictor measure to the regression model (AFQT + Predictor – AFQT Only). Estimates in bold were statistically significant, $p < .05$ (two-tailed). The numbers in brackets after the title of the predictor measure indicate the number of scale scores that the measure contributed to the regression model. The WPA yields six dimension and 14 facet scores. Pairwise deletion was used to account for missing data.

Table 5.4. Incremental Validity Estimates and Predictive Validity Estimates for Experimental Predictors over the AFQT for Predicting Retention-Based Criteria (Dichotomous Criteria)

6-Month Attrition	*n*	AFQT Only	AFQT + Predictor	Δ*R*
Predictor				
AO [1]	4,170	**.045**	**.100**	**.065**
AIM [6]	2,401	**.055**	**.184**	**.129**
TAPAS [12]	2,388	.045	**.197**	**.152**
PSJT [1]	1,618	.063	**.105**	**.042**
RBI [14]	3,442	**.055**	**.212**	**.157**
AKA [6]	4,124	**.045**	**.100**	**.065**
WPA Dimensions [6]	4,096	.045	**.134**	**.089**
WPA Facets [14]	4,093	.045	**.161**	**.116**

Note. The effect sizes reflect Nagelkerke's *R*. Estimates in bold were statistically significant, $p < .05$ (two-tailed). The numbers in brackets after the title of the predictor measure indicate the number of scale scores that the measure contributed to the regression model. The WPA yields six dimension and 14 facet scores. Listwise deletion was used to account for missing data.

With the exception of the AO subtest and the PSJT, the experimental predictor measures exhibited significant potential to increment prediction over the AFQT for career intentions. The validity of the AFQT in predicting career intentions was fairly low (*R* = .05). The RBI (Δ*R* = .27), TAPAS (Δ*R* = .25), and AIM (Δ*R* = .21) again demonstrated the highest levels of incremental validity among experimental predictors. The WPA facets (Δ*R* = .18), WPA dimensions (Δ*R* = .15), and AKA (Δ*R* = .11) also demonstrated substantial incremental validity.

Findings on attrition cognitions were similar to what we found for career intentions and other retention-related criteria. With the exception of the AO and PSJT, the experimental predictor measures exhibited significant potential to increment prediction over the AFQT. The predictive validity of the AFQT for predicting attrition cognitions was low (*R* = .05). Same as career intentions, the RBI (Δ*R* = .26) emerged as the experimental predictor evidencing the highest incremental validity, followed by the TAPAS (Δ*R* = .19), AIM (Δ*R* = .16), WPA facets (Δ*R* = .16), WPA dimensions (Δ*R* = .12), and AKA (Δ*R* = .11).

Finally, the prediction of 6-month attrition was significantly enhanced by all of the experimental predictors. The AFQT exhibited a low level of predictive validity (*R* = .045 – .063). Among the experimental predictors, the RBI (Δ*R* = .16) and TAPAS (Δ*R* = .15) evidenced the greatest incremental validity, followed by AIM (Δ*R* = .13), WPA facets (Δ*R* = .12), WPA dimensions (Δ*R* = .09), AO (Δ*R* = .07), and the PSJT (Δ*R* = .04).

To summarize, compared to the findings from performance-related criteria, the experimental predictors consistently showed greater potential to increment the AFQT in predicting retention-related criteria. Among the experimental predictor measures, the RBI consistently emerged as the predictor measure evidencing the most potential to increment the AFQT in predicting retention-related criteria. The TAPAS and the AIM were also strong predictors of retention-related criteria and demonstrated high incremental validity. The next tier consists of WPA facets, WPA dimensions, and AKA, which were fairly strong predictors and evidenced

moderate levels of incremental validity. Finally, AO and PSJT exhibited comparatively little incremental validity in predicting retention-related criteria.

Subgroup Differences on the Experimental Predictors

The table in Appendix D summarizes subgroup score differences for all eight experimental predictors and the AFQT. The results presented in this table evidence the following:

With a few exceptions, female-male differences on the experimental predictors were generally small and comparable to the AFQT. Across the experimental predictors, the absolute standardized mean difference or average absolute difference (or *d*) ranged from .12 SD (AKA) to .36 SD (PSJT). The AO subtest, AIM, TAPAS, and AKA all evidenced female-male mean differences, on average, that were somewhat smaller or comparable to the AFQT (absolute d = .17), whereas the PSJT, RBI, and the WPA (dimensions and facets) exhibited mean differences, on average, that were roughly double those of the AFQT (absolute average d = .26 to .36). However, in most cases, these mean score differences were such that female Soldiers scored higher, on average, than their male counterparts. The most notable exception to this trend were scales measuring physically-oriented attributes (e.g., the RBI's Fitness Motivation scale, d = -.72; the WPA Realistic Interest dimension scale, d = -.87; the WPA Mechanical and Physical facet scales, with *d*'s of -.83 and -.60, respectively).

Comparatively speaking, Black-White differences on the experimental predictors, excluding AO, were consistently smaller than that observed on the AFQT. Black Soldiers scored .56 SD lower than White Soldiers, on average, on the AFQT. Conversely, the absolute standardized mean difference or average absolute difference (or *d*) on the experimental predictors, excluding AO, ranged from .08 SD (TAPAS) to .41 SD (WPA dimensions). The AIM, TAPAS, PSJT, RBI, and AKA all exhibited average absolute mean differences (or *d*'s) around or less than .20 SD. The WPA dimensions and facets evidenced average mean differences of .41 and .33 SD, respectively. However, when examining the individual scales constituting these measures, most of the mean differences were such that Black Soldiers scored higher, on average, than White Soldiers. The only notable exception is the WPA Realistic Interest dimension scale, d = -.52. Consistent with the AFQT, Black Soldiers scored .59 SD lower than White Soldiers, on average, on the AO subtest.

Hispanic—Non-Hispanic differences on the experimental predictors were consistently lower than that observed on the AFQT. Hispanic Soldiers scored .40 SD lower than non-Hispanic White Soldiers, on average, on the AFQT. This was in contrast to the experimental predictors where the absolute standardized mean difference or average absolute difference (or *d*) on the experimental predictors ranged from .05 SD (AKA) to .26 SD (WPA dimensions). Further, excluding the AO subtest and the PSJT, these mean differences were such that Hispanic Soldiers generally scored higher, on average, than non-Hispanic Soldiers. Unlike the mean differences observed for gender and race, this trend held even at the individual scale level. Across the experimental predictors, there were no individual scales where Hispanic Soldiers scored more than .11 SD lower than non-Hispanic Soldiers.

In sum, the experimental predictors generally exhibited small subgroup differences and differences that were lower, on average, than that observed on the AFQT. Across the different experimental predictors, the observed subgroup differences generally were smaller, on average, than those found on the AFQT, particularly for race and ethnicity. In many instances, the absolute value of the mean differences on the experimental predictors was about half the size of that observed on the AFQT. The direction of these differences was such that minority group members tended to score higher, on average, than majority group members. This trend generally held even at the individual scale level, with a few exceptions of the scales measuring physically-oriented attributes (e.g., the RBI's Fitness Motivation scale, the WPA Realistic Interest dimension scale, the WPA Mechanical and Physical facet scales). Based on the available data, these exceptions likely reflect substantive subgroup differences in those attributes and not the measure's content.

CHAPTER 6: SUMMARY AND CONCLUSIONS

Michael J. Ingerick (HumRRO)

The Army Class longitudinal validation research is designed to provide evidence about the usefulness of several potential measures that could be used to supplement the ASVAB for pre-enlistment screening and classification. This report has described the experimental predictor measures and how they were administered to roughly 11,000 new Soldiers during their first few days in the Army. This report has also described the administration of performance and attitudinal criterion measures at the end of training to over 2,000 Soldiers in six target MOS. These data were the basis for the first set of criterion-related validation analyses for the longitudinal sample. The validation analyses focused on (a) the question of incremental validity over the current primary pre-enlistment screen, the AFQT, and (b) subgroup differences in predictor scores.

Summary of Main Findings

Incremental Validity

In regards to incremental validity, the results of our analyses indicated:

- *The experimental predictors consistently evidenced the potential to increment the AFQT in predicting performance-related criteria, but more so for the behaviorally-based (i.e., what a Soldier does) than the knowledge-based (i.e., what a Soldier knows) criteria.* Overall, the experimental predictors yielded incremental validity estimates (ΔR's) that ranged from .01 to .04 for a knowledge-based criterion (a less than 10% gain over the AFQT), but estimates of upwards of .35 on the more behaviorally-based criteria (a 648% gain over the AFQT). Among the experimental predictors, the RBI, the TAPAS, and the AIM, followed by the WPA, generally evidenced the greatest potential for incrementing the AFQT in predicting Soldier performance during training.

- *The experimental predictors demonstrated substantial gains over the AFQT for predicting retention-related criteria, including early attrition.* The experimental predictors evidenced incremental validity estimates typically in the .10s and as high as .38 (an 800%+ gain over the AFQT) for predicting Soldier attitudes predictive of retention. For predicting early attrition, the experimental predictors bested the AFQT by 66.7% (PSJT) to 285.5% (RBI). Across the retention-related criteria, the RBI generally emerged as the measure demonstrating the greatest gains over the AFQT, followed by the TAPAS, the AIM, and the WPA.

Subgroup Differences

With respect to subgroup differences, our analyses demonstrated:

- *The experimental predictors generally exhibited subgroup differences (for gender, race, and ethnicity) that were lower, on average, than that observed on the AFQT.*

Across the different experimental predictors, the observed subgroup differences, as measured by standardized mean differences (or *d*), were generally smaller, on average, than those found on the AFQT. Specifically, the experimental predictors evidenced subgroup differences that were about half the size of that observed on the AFQT. This finding was particularly true for race and ethnicity, where the average absolute mean differences for the experimental predictors, excluding AO, were upwards of 88% lower than the differences observed on the AFQT (26.8% to 75.0% lower for race; 35.0% to 87.5% lower for ethnicity).

- *Where there were sizeable subgroup differences, their direction tended to be such that minority group members scored higher, on average, than majority group members*. This finding generally held even at the individual scale level, with a few exceptions. Those exceptions were for scales measuring physically-oriented attributes (e.g., the RBI's Fitness Motivation scale, the WPA Realistic Interest dimension scale, the WPA Mechanical and Physical facet scales) where one might expect gender differences.

Limitations and Issues

Comparing Results from the Army Class Longitudinal Validation to the Concurrent Validation

Overall, the pattern of results from the longitudinal validation was comparable to those from the concurrent validation (Ingerick et al., 2009), although the (observed) incremental validity estimates were generally higher in the concurrent validation. However, there are several substantive differences between the two research efforts, excluding differences in sample size, which make a direct comparison inadvisable. Chief among these differences are those pertaining to (a) the research design used in the two efforts (i.e., longitudinal versus concurrent), (b) the characteristics of the Soldiers sampled (i.e., entry-level Soldiers versus incumbent Soldiers in the concurrent validation), and (c) the time in a Soldier's career at which the criterion measures were administered (i.e., at the end-of-training in the longitudinal validation research versus in unit for the concurrent validation research).

Generalizabilty of Findings to an Operational Setting

One of the strengths of the current research effort was the collection of predictor data from entry-level Soldiers at the reception battalions. Doing so enabled us to collect predictor data from Soldiers at an early point in their Army career that was as close to an operational applicant setting as we could get. Although the current research is informative, there are substantive differences between the two settings that could limit the generalizability of these findings to an actual applicant context. Chief among these is that respondents in an operational applicant setting are likely to have a greater motivation to fake or otherwise misrepresent themselves on the experimental predictor measures than in the current research.

Another issue potentially limiting the generalizability of the current findings pertains to the characteristics of Soldiers in the longitudinal validation sample. About half of the predictor sample (about 48%) were Soldiers in our six target MOS. Further, non-administrative training

criterion data were only collected on these Soldiers. Accordingly, the reported findings might not generalize to Soldiers in other MOS. In addition, not all of the predictor measures were administered throughout the predictor data collection, although every effort was made to collect predictor data from large samples of Soldiers throughout the calendar year. For example, the TAPAS and AIM were only administered during the Phase 2 data collections. As a result, our data on these measures was limited to Soldiers who participated at a specific period in the calendar year and might not be fully representative of the Army accession population as a whole.

Future Research

Future research will proceed along two lines. The first will be a continuation of the Army Class longitudinal validation research program and will involve collecting in-unit criterion data, on both performance and retention-related criteria. This will allow examination of the potential of the experimental predictor measures to predict Soldier performance and retention post-training using a longitudinal (as opposed to concurrent) research design. The planned two rounds of in-unit criterion data collection will include all Soldiers in the longitudinal sample (not just those in the six target MOS) and will hopefully permit more extensive analyses to examine the classification potential of the experimental predictor measures.

REFERENCES

Campbell, J.P., & Knapp, D.J. (Eds.) (2001). *Exploring the limits in personnel selection and classification*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Campbell, J.P., McCloy, R.A., McPhail, S.M., Pearlman, K., Peterson, N.G., Rounds, J., & Ingerick, M. (2007). *U.S. Army Classification Research Panel: Conclusions and recommendations on classification research strategies* (Study Report 2007-05). Arlington, VA: U. S. Army Research Institute for the Behavioral and Social Sciences.

Campbell, J.P., McCloy, R.A., Oppler, S.H., & Sager, C.E. (1993). A theory of performance. In N. Schmitt, W.C. Borman, & Associates (Eds.), *Personnel selection in organizations* (pp. 35-70). San Francisco: Jossey-Bass Publishers.

Collins, M., Le, H., & Schantz, L. (2005). Job knowledge criterion tests. In D.J. Knapp & T.R. Tremble (Eds.), *Development of experimental Army enlisted personnel selection and classification tests and job performance criteria* (Technical Report 1168) (pp. 49-58). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Department of Defense (2008). Office of the Under Secretary of Defense, Personnel and Readiness. *Population representation in the military services: Fiscal Year 2006* (FR-08-27). Washington, DC: Author. http://www.defenselink.mil/prhome/PopRep_FY06/

Drasgow, F. , Stark, S., & Chernyshenko, O.S. (November, 2006). *Toward the next generation of personality assessment systems to support personnel selection and classification decisions*. Paper presented at the 48th annual conference of the International Military Testing Association, Canada.

Holland, J.L. (1997). *Making vocational choices: A theory of vocational personalities and work environments* (3rd ed.). Odessa, FL: Psychological Assessment Resources, Inc.

Hoffman, R.R., Muraca, S.T., Heffner, T.S., Hendricks, R., & Hunter, A.E. (2009). *Selection for accelerated basic combat training* (Technical Report 1241). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Ingerick, M., Diaz, T., & Putka, D. (2009). *Investigations into Army enlisted classification systems: Concurrent validation report* (Technical Report 1244). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Keenan, P.A., Russell, T.L., Le., H., Katkowski, D., & Knapp, D.J. (2005). Performance rating scales. In D.J. Knapp & T.R. Tremble (Eds.), *Development of experimental Army enlisted personnel selection and classification tests and job performance criteria* (pp. 21-48) (Technical Report 1168). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Kilcullen, R.N., Mael, F.A., Goodwin, G.F., & Zazanis, M.M. (1999). Predicting U.S. Army Special Forces field performance. *Human Performance in Extreme Environments, 4(1),* 53-63.

Kilcullen, R.N., Putka, D.J., McCloy, R.A., & Van Iddekinge, C.H. (2005). Development of the Rational Biodata Inventory. In D.J. Knapp, C.E. Sager, & T.R. Tremble (Eds.), *Development of experimental Army enlisted personnel selection and classification tests and job performance criteria* (pp. 105-116) (Technical Report 1168). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Kilcullen, R.N., White, L.A., Sanders, M. & Hazlett, G. (2003). *The Assessment of Right Conduct (ARC) administrator's manual* (Research Note 2003-09). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Knapp, D.J., & Campbell, R.C. (Eds). (2006). *Army enlisted personnel competency assessment program: Phase II report* (Technical Report 1174). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Knapp, D. J., & Heffner, T. S. (Eds.). (2009). *Expanded Enlistment Eligibility Metrics (EEEM): Recommendations on a non-cognitive screen for new soldier selection* (Technical Report XXXX). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Knapp, D.J., McCloy, R.A., & Heffner, T.S. (Eds.) (2004). *Validation of measures designed to maximize 21st-century Army NCO performance* (Technical Report 1145). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Knapp, D. J., Sager, C. E., & Tremble, T. R. (Eds.) (2005). *Development of experimental Army enlisted personnel selection and classification tests and job performance criteria* (Technical Report 1168). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Knapp, D.J., & Tremble, T.R. (Eds) (2007). *Concurrent validation of experimental Army enlisted personnel selection and classification measures* (Technical Report 1205). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Moriarty, K.O., Campbell, R.C., Heffner, T.S., & Knapp, D.J. (2009). *Investigations into Army enlisted classification systems (Army Class): Reclassification test and criterion development report* (FR 08-54). Alexandria, VA: Human Resources Research Organization.

Peterson, N.G., Russell, T.L., Hallam, G., Hough, L.M., Owens-Kurtz, C., Gialluca, K., & Kerwin, K. (1992). Analysis of the experimental predictor battery: LV sample. In J. P.Campbell & L.M. Zook (Eds.), *Building and retaining the career force: New procedures for accessing and assigning Army enlisted personnel-Annual report, 1990 fiscal year* (Technical Report 952) (pp. 73-199). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Putka, D.J., & Van Iddekinge, C.H. (2007). Work Preferences Survey. In D.J. Knapp & T.R. Tremble (Eds.), *Concurrent validation of experimental Army enlisted personnel selection and classification measures* (Technical Report 1205). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Russell, T.L., Peterson, N.G., Rosse, R.L., Hatten, J.L.T., McHenry, J. J., & Houston, J.S. (2001). The measurement of cognitive, perceptual and psychomotor abilities. In J.P. Campbell & D.J. Knapp (Eds.), *Exploring the limits in personnel selection and classification* (pp. 71-110). Mahwah, NJ: Lawrence Erlbaum Inc.

Russell, T.L., Reynolds, D.H., & Campbell, J.P. (Eds.) (1994). *Building a joint service classification research roadmap: Individual differences measurement* (AL/HR-TP-1994-0009). Brooks AFB TX: Armstrong Laboratory.

Stark, S. (2002). *A new IRT approach to test construction and scoring designed to reduce the effects of faking in personality assessment* [Doctoral Dissertation]. University of Illinois at Urbana-Champaign.

Stark, S., Drasgow, F., & Chernyshenko, O.S. (October, 2008). *Update on Tailored Adaptive Personality Assessment System (TAPAS): The next generation of personality assessment systems to support personnel selection and classification decisions*. Paper presented at the 50th annual conference of the International Military Testing Association, Amsterdam, Netherlands.

Van Iddekinge, C.H., Putka, D.J., & Sager, C.E. (2005). Attitudinal criteria. In D.J. Knapp & T.R. Tremble (Eds.), *Development of experimental Army enlisted personnel selection and classification tests and job performance criteria* (pp. 89-104) (Technical Report 1168). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Viswesvaran, C., Ones, D.S., & Schmidt, F.L. (1996). Comparative analysis of the reliability of job performance ratings. *Journal of Applied Psychology, 81*, 557-574.

Waugh, G.W., & Russell, T.L. (2005). Predictor situational judgment test. In D.J. Knapp & T.R. Tremble (Eds.), *Development of experimental Army enlisted personnel selection and classification tests and job performance criteria* (pp. 235-154) (Technical Report 1168). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

White, L.A., & Young, M.C. (1998, August). *Development and validation of the Assessment of Individual Motivation (AIM)*. Paper presented at the annual meeting of the American Psychological Association, San Francisco, CA.

White, L.A., Young, M.C., & Rumsey, M.G. (2001). ABLE implementation issues and related research. In J.P. Campbell & D.J. Knapp (Eds.), *Exploring the limits of personnel selection and classification* (pp. 525-558). Mahwah, New Jersey: Lawrence Erlbaum Associates.

APPENDIX A
DESCRIPTIVE STATISTICS AND SCORE INTERCORRELATIONS FOR SELECTED CRITERION MEASURES

Table A.1. Descriptive Statistics and Reliability Estimates for the Army-Wide (AW) and MOS-Specific Performance Rating Scales (PRS)

	11B Infantryman		19K Armor Crewmen		31B Military Police		63B Light Wheel Vehicle Mechanic		68W Health Care Specialist		88M Motor Transport Operator		Total				
Composite/Scale	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*α*	*ICC(A,1)*	*ICC(A,k)*
AW PRS																	
Effort Composite	3.56	.79	3.54	.71	3.50	.70	3.50	.75	3.81	.68	3.63	.61	3.56	.74	.90	.30	.63
Physical Fitness & Bearing Composite	3.96	.73	3.79	.68	3.92	.74	3.83	.68	4.01	.57	3.90	.61	3.90	.71	.87	.31	.63
Personal Discipline Composite	3.95	.72	3.76	.63	3.73	.73	3.66	.81	4.11	.53	3.72	.66	3.82	.71	.91	.28	.61
Commitment & Adjustment Composite	3.93	.73	3.69	.70	3.76	.69	3.71	.76	4.05	.59	3.63	.70	3.80	.72	.86	.23	.54
Support for Peers Composite	3.83	.70	3.75	.61	3.77	.63	3.71	.68	4.05	.45	3.75	.63	3.79	.65	.86	.17	.45
Peer Leadership Composite	3.48	.86	3.40	.76	3.32	.77	3.50	.80	3.74	.71	3.38	.73	3.43	.80	.90	.26	.58
Common Warrior Tasks KS Scale	3.91	.76	3.77	.69	3.93	.70	3.86	.71	4.06	.52	3.83	.67	3.89	.71	n/a	.20	.49
MOS Qualification KS Scale	3.97	.76	3.82	.64	3.96	.66	4.02	.74	4.06	.58	3.98	.60	3.95	.69	n/a	.17	.45
MOS-Specific PRS Composite[a]	5.14	.92	4.68	.75	5.03	.74	5.29	1.10	5.38	.74	5.24	.76	5.13	.84	.93	.18	.37

Note. n = 2,229-2,274; 11B Infantryman n = 644-659; 19K Armor Crewmen n = 469-470; 31B Military Police n = 700-715; 63B Light Wheel Vehicle Mechanic n = 214-222; 68W Health Care Specialist n = 129-136; 88M Motor Transport Operator n = 72-73. α = coefficient alpha. n/a = single-item measure. *ICC(A,1)* = intraclass correlation coefficient assuming a single rater. *ICC(A,k)* = intraclass correlation coefficient assuming multiple (or *k*) raters. The AW PRS scales range from 1 – 5; the MOS-Specific PRS Composite ranges from 1 – 7.

[a] The mean, standard deviation, and reliability estimates for the total sample are unit-weighted averages of the estimates for the individual MOS; *α* (11B = .96, 19K = .93, 31B = .94, 63B = .96, 68W = .91, 88M = .88), *ICC(A,1)* (11B = .17, 19K = .17, 31B = .23, 63B = .25, 68W = .11, 88M = .16), and *ICC(A,k)* (11B = .39, 19K = .37, 31B = .48, 63B = .44, 68W = .25, 88M = .30). Ratings include both peers and supervisors.

Table A.2. Intercorrelations among Army-Wide (AW) and MOS-Specific PRS

	Composite/Scale	1	2	3	4	5	6	7	8
1	AW Effort Composite								
2	AW Physical Fitness & Bearing Composite	.74							
3	AW Personal Discipline Composite	.79	.64						
4	AW Commitment & Adjustment Composite	.78	.75	.81					
5	AW Support for Peers Composite	.74	.63	.80	.78				
6	AW Peer Leadership Composite	.76	.70	.68	.75	.73			
7	AW Common Warrior Tasks KS Scale	.73	.76	.67	.78	.68	.73		
8	MOS Qualification KS Scale	.69	.70	.64	.74	.65	.68	.80	
9	MOS-Specific PRS Composite - Total	.63	.60	.58	.65	.59	.64	.66	.67
9a	MOS-Specific PRS Composite - 11B	.69	.67	.66	.73	.67	.68	.76	.74
9b	MOS-Specific PRS Composite - 19K	.54	.48	.53	.55	.58	.54	.56	.56
9c	MOS-Specific PRS Composite - 31B	.70	.65	.57	.69	.58	.72	.71	.72
9d	MOS-Specific PRS Composite - 63B	.68	.63	.60	.64	.62	.60	.61	.63
9e	MOS-Specific PRS Composite - 68W	.52	.40	.48	.54	.41	.55	.37	.41
9f	MOS-Specific PRS Composite - 88M	.51	.56	.55	.49	.57	.61	.64	.49

Note. n = 73-2,277. The correlations between the MOS-specific composite ratings and the AW composites/scales for each MOS are presented in rows 9a through 9d. 11B Infantryman n = 642; 19K Armor Crewman n = 469; 31B Military Police n = 703; 63B Light Wheel Vehicle Mechanic n = 214; 68W Health Care Specialist n = 129; and 88M Motor Transport Operator n = 73. All correlations are statistically significant, $p < .05$ (two-tailed).

Table A.3. Descriptive Statistics and Reliability Estimates for the Army Life Questionnaire (ALQ) Scales by MOS

	11B Infantryman		19K Armor Crewman		31B Military Police		63B Light Wheel Vehicle Mechanic	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Commitment and Retention-Related Attitudes								
Attrition Cognitions	4.47	.65	4.36	.65	4.28	.72	4.29	.76
Career Intentions	3.35	1.04	3.18	1.00	3.02	1.01	3.14	1.00
Army Fit	4.15	.57	4.07	.56	4.01	.63	3.93	.68
MOS Fit	3.86	.85	3.36	.87	3.76	.83	3.57	.89
Normative Commitment	4.21	.67	4.10	.69	3.94	.79	4.04	.78
Affective Commitment	3.99	.65	3.97	.60	3.79	.67	3.79	.69
Initial Entry Training (IET) Performance and Adjustment								
Adjustment to Army Life	3.73	.68	3.62	.68	3.70	.69	3.70	.70
Number of Disciplinary Incidents	.38	.79	.46	.87	.58	.98	.53	.92
Last APFT Score	248.53	31.02	238.71	28.32	244.61	33.48	243.81	32.89
Number of IET Achievements	.59	.70	.55	.71	.42	.59	.48	.56
Number of IET Failures	.36	.58	.36	.59	.43	.63	.50	.67
Self-Rated AIT/OSUT Performance								
Physical Fitness	3.13	1.14	3.05	1.14	2.96	1.09	2.93	1.10
Discipline	3.49	1.13	3.46	1.12	3.25	1.14	3.33	1.19
Field Exercises	3.44	.99	3.36	.98	3.08	.98	3.22	1.05
Classroom & Instructional Modules	2.78	1.07	3.15	1.11	3.00	1.01	3.32	1.14
Self-Ranked AIT/OSUT Performance								
Physical Fitness	2.39	1.10	2.51	1.13	2.45	1.17	2.48	1.17
Discipline	2.06	.96	1.92	.97	2.06	.95	2.15	.96
Field Exercises	2.08	.93	2.42	.98	2.44	1.01	2.70	1.07
Classroom & Instructional Modules	3.46	.86	3.15	1.03	3.05	1.10	2.67	1.18

Table A.3 (continued)

	68W Health Care Specialist		88M Motor Transport Operator		Total Sample		
	M	*SD*	*M*	*SD*	*M*	*SD*	α
Commitment and Retention-Related Attitudes							
Attrition Cognitions	4.22	.79	4.33	.75	4.35	.70	.80
Career Intentions	3.01	1.13	3.36	1.06	3.17	1.03	.94
Army Fit	3.81	.79	4.05	.67	4.04	.62	.82
MOS Fit	3.90	.87	3.17	.95	3.68	.88	.93
Normative Commitment	3.91	.91	3.98	.77	4.06	.75	.79
Affective Commitment	3.53	.92	3.84	.66	3.87	.68	.87
Initial Entry Training (IET) Performance and Adjustment							
Adjustment to Army Life	3.70	.73	3.59	.73	3.69	.69	.82
Number of Disciplinary Incidents	.45	.84	.84	1.42	.49	.91	n/a
Last APFT Score	257.95	29.08	242.10	33.17	245.18	31.73	n/a
Number of IET Achievements	.39	.51	.54	.61	.51	.65	n/a
Number of IET Failures	.54	.69	.49	.68	.41	.62	n/a
Self-Rated AIT/OSUT Performance							
Physical Fitness	3.19	1.19	3.04	1.11	3.04	1.12	n/a
Discipline	3.54	1.22	3.67	1.02	3.40	1.14	n/a
Field Exercises	3.22	.98	3.53	.99	3.28	1.00	n/a
Classroom & Instructional Modules	3.20	1.10	3.33	1.21	3.02	1.09	n/a
Self-Ranked AIT/OSUT Performance							
Physical Fitness	2.53	1.16	2.57	1.11	2.46	1.14	n/a
Discipline	2.29	1.10	2.00	1.00	2.05	.97	n/a
Field Exercises	2.67	1.01	2.43	.98	2.37	1.01	n/a
Classroom & Instructional Modules	2.51	1.18	2.99	1.17	3.12	1.07	n/a

Note. *n* = 2,191-2,214; 11B Infantryman *n* = 640-665; 19K Armor Crewman *n* = 453-463; 31B Military Police *n* = 672-684; 63B Light Wheel Vehicle Mechanic *n* = 211-215; 68W Health Care Specialist *n* = 133; 88M Motor Transport Operator *n* = 68-70. APFT = Army Physical Fitness Test. IET = Initial Entry Training. α = coefficient alpha. ALQ scale scores range from 1 – 5 except for the following: (a) Number of Disciplinary Incidents (0 – 7), (b) Last APFT Score (free response item, Min = 62, Max = 300), (c) Number of IET Achievements (0 – 2), (d) Number of IET Failures (0 – 3), (e) Soldiers' self-rated AIT/OSUT performance (1 – 4; 1 = *Below Average [Bottom 30%]* to 4 = *Truly Exceptional [Top 5%]*), (f) Soldiers' self-ranked AIT/OSUT performance (1 – 4, where 1 = *Strongest Area of Performance* and 4 = *Weakest Area of Performance.*

Table A.4. Intercorrelations among ALQ Scale Scores

Scale		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	Attrition Cognitions																		
2	Career Intentions	**.53**																	
3	Army Fit	**.72**	**.55**																
4	MOS Fit	**.41**	**.32**	**.49**															
5	Normative Commitment	**.75**	**.52**	**.68**	**.41**														
6	Affective Commitment	**.65**	**.58**	**.79**	**.48**	**.68**													
7	Adjustment to Army Life	**.52**	**.37**	**.61**	**.35**	**.41**	**.41**												
8	# of Disciplinary Incidents	**-.21**	**-.07**	**-.23**	**-.12**	**-.16**	**-.14**	**-.25**											
9	Last APFT Score	**.11**	**.05**	**.16**	**.09**	**.09**	**.05**	**.26**	**-.14**										
10	# of IET Achievements	**.07**	**.09**	**.14**	**.05**	**.05**	**.09**	**.16**	**-.08**	**.27**									
11	# of IET Failures	**-.15**	**-.08**	**-.16**	**-.10**	**-.11**	**-.09**	**-.25**	**.19**	**-.29**	**-.17**								
12	Self-Rating (PHYS)	**.07**	.04	**.14**	**.06**	**.06**	**.05**	**.22**	**-.08**	**.62**	**.31**	**-.26**							
13	Self-Rating (DISC)	**.18**	**.08**	**.24**	**.09**	**.16**	**.17**	**.23**	**-.21**	**.12**	**.19**	**-.10**	**.25**						
14	Self-Rating (FX)	**.17**	**.16**	**.23**	**.16**	**.14**	**.21**	**.21**	.00	**.10**	**.20**	**-.16**	**.25**	**.25**					
15	Self-Rating (INST)	**.09**	.02	**.09**	.03	**.08**	**.06**	**.14**	-.03	-.01	**.08**	**-.07**	**.08**	**.24**	**.30**				
16	Self-Ranking (PHYS)	.00	.01	-.04	-.01	.03	.03	**-.12**	.03	**-.53**	**-.21**	**.19**	**-.55**	**.05**	.02	**.19**			
17	Self-Ranking (DISC)	**-.10**	-.03	**-.14**	.00	**-.08**	**-.11**	-.04	**.13**	**.14**	-.03	**-.07**	**.14**	**-.36**	**.12**	**.07**	**-.32**		
18	Self-Ranking (FX)	-.02	**-.07**	.00	**-.06**	-.03	**-.06**	.02	**-.07**	**.19**	**.09**	-.01	**.18**	**.15**	**-.31**	**.08**	**-.30**	**-.36**	
19	Self-Ranking (INST)	**.11**	**.08**	**.17**	**.07**	**.07**	**.12**	**.15**	**-.09**	**.26**	**.17**	**-.13**	**.29**	**.12**	**.16**	**-.34**	**-.49**	**-.23**	**-.29**

Note. n = 2,173-2,216. Statistically significant correlations are bolded, $p < .05$ (two-tailed). PHYS = Physical Fitness, DISC = Discipline, FX = Field Exercises, INST = Classroom and Instructional Modules, APFT = Army Physical Fitness Test, IET = Initial Entry Training.

APPENDIX B
DESCRIPTIVE STATISTICS AND SCORE INTERCORRELATIONS FOR SELECTED PREDICTOR MEASURES

Table B.1. Descriptive Statistics for the Armed Services Vocational Aptitude Battery (ASVAB) Subtests and Armed Forces Qualification Test (AFQT)

Scale	*M*	*SD*
ASVAB Subtests		
General Science (GS)	51.34	7.36
Arithmetic Reasoning (AR)	51.82	6.29
Word Knowledge (WK)	49.94	5.97
Paragraph Comprehension (PC)	51.47	5.09
Math Knowledge (MK)	52.17	6.30
Electronics Information (EI)	52.04	7.79
Auto and Shop Information (AS)	50.76	8.56
Mechanical Comprehension (MC)	53.18	7.62
Assembling Objects (AO)	54.88	7.95
AFQT	56.13	19.31

Note. n = 9,467-10,736. Subtest and composite scores are percentiles.

Table B.2. Intercorrelations among ASVAB Subtest and AFQT Scores

Scale		1	2	3	4	5	6	7	8	9
1	General Science (GS)									
2	Arithmetic Reasoning (AR)	.39								
3	Word Knowledge (WK)	.61	.25							
4	Paragraph Comprehension (PC)	.43	.28	.43						
5	Math Knowledge (MK)	.28	.56	.09	.15					
6	Electronics Information (EI)	.57	.36	.43	.32	.16				
7	Auto and Shop Information (AS)	.42	.25	.29	.20	-.03	.58			
8	Mechanical Comprehension (MC)	.52	.45	.36	.30	.24	.58	.57		
9	Assembling Objects (AO)	.30	.39	.16	.19	.32	.31	.23	.49	
10	AFQT	.66	.76	.70	.62	.65	.49	.28	.52	.41

Note. n = 9,084 – 10,633. All correlations are statistically significant, $p < .05$ (two-tailed).

Table B.3. Descriptive Statistics and Reliability Estimates for Assessment of Individual Motivation (AIM) Scales

Scale	*M*	*SD*	α
Adjustment	1.26	.29	.74
Agreeableness	1.26	.27	.70
Dependability	1.26	.28	.77
Leadership	1.20	.28	.76
Physical Conditioning	1.19	.34	.78
Work Orientation	1.20	.29	.74
Validity Scale	.15	.16	n/a

Note. n = 4,707 – 4,939. α = coefficient alpha. AIM scales scores range from 0 – 2 except for the Validity scale, which ranges from 0 – 1.

Table B.4. Intercorrelations among AIM Scales

Scale		1	2	3	4	5	6
1	Adjustment						
2	Agreeableness	**.63**					
3	Dependability	**.52**	**.52**				
4	Leadership	**.29**	**.17**	**.37**			
5	Physical Conditioning	**.30**	**.29**	**.31**	**.24**		
6	Work Orientation	**.40**	**.32**	**.34**	**.57**	**.54**	
7	Validity Scale	**.11**	**.09**	**.08**	**.04**	.02	**.13**

Note. n = 4,696 – 4,939. Statistically significant correlations are bolded, $p < .05$ (two-tailed).

Table B.5. Descriptive Statistics for Tailored Adaptive Personality Assessment System (TAPAS-95s) Scales

Scale	Items	*M*	*SD*
Achievement	16	.17	.64
Curiosity	13	-.08	.79
Non-Delinquency	17	.09	.65
Dominance	17	-.15	.61
Even-Temper	13	-.46	.76
Attention-Seeking	14	-.14	.79
Intellectual Efficiency	14	-.19	.64
Order	13	-.04	.64
Physical Conditioning	17	.12	.71
Tolerance	13	-.43	.67
Cooperation/Trust	17	-.30	.86
Optimism	15	-.07	.59

Note. n = 4,637. Scores have a theoretical distribution of approximately -3 to +3.

Table B.6. Intercorrelations among TAPAS-95s Scales

	Scale	1	2	3	4	5	6	7	8	9	10	11
1	Achievement											
2	Curiosity	**.21**										
3	Non-Delinquency	**.17**	**.12**									
4	Dominance	**.15**	**.14**	.02								
5	Even-Temper	**.06**	**.22**	**.12**	**-.05**							
6	Attention-Seeking	**-.11**	**-.11**	**-.37**	**.13**	**-.12**						
7	Intellectual Efficiency	**.16**	**.34**	**.03**	**.15**	**.14**	**-.06**					
8	Order	**.19**	**.05**	**.15**	**.07**	-.02	**-.07**	**.07**				
9	Physical Conditioning	**.19**	**.04**	**-.09**	**.06**	-.01	**.10**	.02	**.05**			
10	Tolerance	**.06**	**.21**	**.06**	**.10**	**.08**	**-.03**	**.15**	**.06**	.01		
11	Cooperation/Trust	.01	**-.05**	**.19**	**-.13**	**.12**	**-.05**	**-.07**	.02	**-.13**	-.01	
12	Optimism	**.06**	**.12**	.03	**.08**	**.22**	**-.03**	**.17**	.00	**.07**	**.09**	**.09**

Note. n = 4,637. Statistically significant correlations are bolded, $p < .05$ (two-tailed).

Table B.7. Descriptive Statistics and Reliability Estimates for Rational Biodata Inventory (RBI) Scale Scores

Scale	Items	*M*	*SD*	α
Peer Leadership	6	3.60	.65	.71
Cognitive Flexibility	8	3.47	.64	.76
Achievement	9	3.54	.58	.70
Fitness Motivation	7	3.30	.68	.73
Interpersonal Skills - Diplomacy	5	3.65	.75	.71
Stress Tolerance	11	3.01	.51	.67
Hostility to Authority	7	2.52	.65	.68
Self-Efficacy	6	4.02	.62	.78
Cultural Tolerance	5	3.75	.73	.69
Internal Locus of Control	8	3.55	.57	.67
Army Affective Commitment	7	3.73	.69	.71
Respect for Authority	4	3.51	.69	.65
Narcissism	6	3.61	.57	.55
Gratitude	3	3.95	.72	.43
Lie Scale	7	0.09	.14	.51
Pure Fitness Motivation [a]	5	3.40	.72	.70

Note. *n* = 8,625-8,626. Items = number of items comprising each final scale. α = coefficient alpha. RBI scale scores range from 1 – 5, except for the Lie scale, which ranges from 0 – 1.

[a] An alternative version of the Fitness Motivation scale with the ability items removed.

Table B.8. Intercorrelations among RBI Scale Scores

	Scale	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	Peer Leadership															
2	Cognitive Flexibility	**.51**														
3	Achievement	**.55**	**.49**													
4	Fitness Motivation	**.29**	**.16**	**.27**												
5	Interpersonal Skills - Diplomacy	**.49**	**.30**	**.38**	**.22**											
6	Stress Tolerance	**.12**	**.14**	**.06**	**.22**	**.24**										
7	Hostility to Authority	**-.10**	**-.18**	**-.25**	**-.05**	**-.18**	**-.37**									
8	Self-Efficacy	**.57**	**.44**	**.56**	**.38**	**.46**	**.24**	**-.19**								
9	Cultural Tolerance	**.35**	**.42**	**.31**	**.13**	**.42**	**.30**	**-.34**	**.40**							
10	Internal Locus of Control	**.31**	**.28**	**.35**	**.21**	**.37**	**.42**	**-.39**	**.45**	**.38**						
11	Army Affective Commitment	**.31**	**.19**	**.29**	**.30**	**.29**	**.22**	**-.20**	**.44**	**.27**	**.34**					
12	Respect for Authority	**.28**	**.29**	**.49**	**.10**	**.20**	-.01	**-.21**	**.30**	**.19**	**.21**	**.19**				
13	Narcissism	**.37**	**.23**	**.34**	**.18**	**.21**	**-.15**	**.15**	**.39**	**.08**	**.10**	**.18**	**.15**			
14	Gratitude	**.27**	**.24**	**.34**	**.12**	**.33**	**.10**	**-.28**	**.35**	**.30**	**.35**	**.24**	**.32**	**.11**		
15	Lie Scale	**.16**	**.15**	**.17**	**.12**	**.12**	**.24**	**-.20**	**.19**	**.20**	**.17**	**.12**	**.09**	**.02**	.01	
16	Pure Fitness Motivation [a]	**.32**	**.20**	**.33**	**.93**	**.24**	**.19**	**-.08**	**.42**	**.17**	**.23**	**.34**	**.14**	**.19**	**.16**	**.13**

Note. n = 8,624-8,626. Statistically significant correlations are bolded, $p < .05$ (two-tailed).

[a] An alternative version of the Fitness Motivation scale with the ability items removed.

Table B.9. Descriptive Statistics and Reliability Estimates for Army Knowledge Assessment (AKA) Scales

Scale	Items	*M*	*SD*	α
Realistic Interests	5	4.05	.61	.76
Investigative Interests	5	3.39	.74	.82
Artistic Interests	5	2.75	.93	.89
Social Interests	5	3.78	.71	.82
Enterprising Interests	5	3.69	.71	.81
Conventional Interests	5	3.93	.69	.84

Note. *n* = 10,048-10,075. Items = number of items comprising each final scale. α = coefficient alpha. AKA scale scores range from 1 – 5.

Table B.10. Intercorrelations among AKA Scales

Scale	1	2	3	4	5
1 Realistic Interests					
2 Investigative Interests	.39				
3 Artistic Interests	.14	.50			
4 Social Interests	.39	.38	.30		
5 Enterprising Interests	.40	.38	.25	.48	
6 Conventional Interests	.44	.29	.10	.45	.52

Note. *n* = 10,044 – 10,074. All correlations are statistically significant, $p < .05$ (two-tailed).

Table B.11. Descriptive Statistics and Reliability Estimates for Work Preferences Assessment (WPA) Dimension and Facet Scores

Scale	Items	*M*	*SD*	α
Realistic Interests (D)	13	3.50	.79	.90
Mechanical (F)	5	3.20	1.05	.90
Physical (F)	7	3.73	.84	.89
Investigative Interests (D)	12	3.28	.65	.85
Critical Thinking (F)	6	3.76	.72	.82
Conduct Research (F)	6	2.79	.77	.76
Artistic Interests (D)	12	2.79	.76	.87
Artistic Activities (F)	8	2.39	.86	.85
Creativity (F)	4	3.59	.86	.82
Social Interests (D)	10	3.60	.65	.83
Work with Others (F)	5	3.81	.71	.77
Help Others (F)	5	3.39	.75	.71
Enterprising Interests (D)	13	3.36	.59	.81
Prestige (F)	5	3.88	.66	.68
Lead Others (F)	4	3.56	.74	.70
High Profile (F)	4	2.52	.88	.72
Conventional Interests (D)	12	3.23	.62	.82
Information Management (F)	6	2.63	.84	.81
Detail Orientation (F)	3	3.88	.78	.73
Clear Procedures (F)	3	3.90	.76	.64

Note. *n* = 9,924-9,926. D = Dimension. F = Facet. Items = number of items comprising each final scale. α = coefficient alpha. WPA scale scores range from 1 – 5.

Table B.12. Intercorrelations among WPA Dimension and Facet Scores

Scale	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1 Realistic Interests (D)																			
2 Mechanical (F)	**.83**																		
3 Physical (F)	**.86**	**.45**																	
4 Investigative Interests (D)	**.16**	**.12**	**.15**																
5 Critical Thinking (F)	**.20**	**.09**	**.25**	**.86**															
6 Conduct Research (F)	**.08**	**.12**	**.02**	**.88**	**.52**														
7 Artistic Interests (D)	**.10**	**.18**	.01	**.42**	**.24**	**.47**													
8 Artistic Activities (F)	**.08**	**.17**	**-.03**	**.31**	**.10**	**.43**	**.94**												
9 Creativity (F)	**.12**	**.13**	**.08**	**.47**	**.43**	**.40**	**.76**	**.50**											
10 Social Interests (D)	**.09**	**-.06**	**.19**	**.54**	**.53**	**.41**	**.29**	**.21**	**.35**										
11 Work with Others (F)	**.20**	.00	**.32**	**.45**	**.51**	**.29**	**.20**	**.11**	**.30**	**.88**									
12 Help Others (F)	**-.03**	**-.10**	**.04**	**.50**	**.43**	**.44**	**.32**	**.26**	**.31**	**.90**	**.58**								
13 Enterprising Interests (D)	**.16**	**.06**	**.19**	**.61**	**.57**	**.50**	**.39**	**.30**	**.42**	**.59**	**.54**	**.51**							
14 Prestige (F)	**.18**	**.05**	**.24**	**.50**	**.55**	**.32**	**.19**	**.08**	**.34**	**.50**	**.50**	**.39**	**.80**						
15 Lead Others (F)	**.21**	**.03**	**.31**	**.48**	**.52**	**.32**	**.24**	**.14**	**.35**	**.59**	**.56**	**.49**	**.81**	**.57**					
16 High Profile (F)	.00	**.07**	**-.07**	**.46**	**.28**	**.51**	**.46**	**.46**	**.30**	**.32**	**.23**	**.34**	**.74**	**.33**	**.37**				
17 Conventional Interests (D)	**.12**	**.12**	**.08**	**.61**	**.53**	**.53**	**.26**	**.23**	**.24**	**.55**	**.48**	**.51**	**.59**	**.47**	**.42**	**.48**			
18 Information Management (F)	**-.05**	**.07**	**-.14**	**.49**	**.31**	**.54**	**.36**	**.37**	**.22**	**.41**	**.28**	**.44**	**.51**	**.28**	**.29**	**.61**	**.85**		
19 Detail Orientation (F)	**.24**	**.13**	**.28**	**.53**	**.61**	**.32**	**.07**	**-.02**	**.23**	**.47**	**.49**	**.36**	**.42**	**.48**	**.39**	**.13**	**.69**	**.30**	
20 Clear Procedures (F)	**.21**	**.11**	**.24**	**.48**	**.54**	**.30**	**.05**	**-.02**	**.18**	**.49**	**.49**	**.39**	**.41**	**.48**	**.37**	**.14**	**.72**	**.33**	**.89**

Note. n = 9,924-9,926. D = Dimension. F = Facet. Statistically significant correlations are bolded, $p < .05$ (two-tailed).

APPENDIX C
SCALE-LEVEL CORRELATIONS BETWEEN SELECTED PREDICTOR AND CRITERION MEASURES

Table C.1. Correlations between Predictor Scale Scores and Selected Performance-Related Criterion Measures

	Criterion Measure/Scale								
Predictor Measure/Scale	MOS-SPEC JKT	MOS-SPEC PRS	EFFORT PRS	PHYS FIT PRS	APFT SCORE	PEERS PRS	LEADER PRS	PER DISC PRS	DISC INC
AFQT	**.44**	**.15**	**.19**	**.09**	**.05**	**.15**	**.15**	**.18**	**-.09**
Assembling Objects (AO)	**.29**	**.14**	**.18**	**.14**	.04	**.14**	**.15**	**.16**	**-.08**
TAPAS-95s									
Achievement	**.09**	**.07**	**.09**	**.07**	**.12**	-.02	.04	.05	-.04
Curiosity	**.14**	.04	.04	.01	.06	.02	.06	.05	-.05
Non-Delinquency	**.08**	.01	.05	-.05	**-.09**	**.08**	.01	**.15**	**-.08**
Dominance	.03	-.00	-.04	-.03	.04	**-.08**	-.02	**-.09**	.03
Even-Temper	**.12**	.03	**.09**	.05	-.02	**.09**	.04	**.15**	**-.08**
Attention-Seeking	-.02	-.04	**-.10**	-.03	.01	**-.11**	-.05	**-.17**	**.10**
Intellectual Efficiency	**.15**	.04	.05	-.02	.05	-.01	.03	.03	.01
Order	-.01	.05	.01	.00	.05	-.01	.02	.02	-.05
Physical Conditioning	-.01	**.09**	**.10**	**.21**	**.30**	.00	**.14**	.02	-.01
Tolerance	.00	.02	.02	.01	.01	.06	.05	.06	-.01
Cooperation/Trust	-.05	-.03	.01	-.02	**-.12**	.04	-.03	**.08**	-.05
Optimism	**.14**	.03	.02	.01	.01	.02	.05	.04	.00
AIM									
Adjustment	**.12**	.07	**.12**	**.09**	**.08**	**.10**	**.12**	**.12**	-.06
Agreeableness	**.08**	**.13**	**.16**	**.11**	.04	**.15**	**.13**	**.16**	**-.10**
Dependability	**.10**	**.12**	**.11**	.06	**.09**	**.09**	**.10**	**.16**	**-.09**
Leadership	**.10**	.07	.07	**.08**	**.21**	.01	**.11**	.03	.01
Physical Conditioning	.05	**.09**	**.16**	**.26**	**.31**	**.13**	**.20**	**.13**	-.07
Work Orientation	**.09**	.05	.06	**.08**	**.19**	-.01	**.11**	-.02	.01
RBI									
Peer Leadership	.01	.03	.03	**.05**	**.13**	.03	**.06**	-.03	-.01
Cognitive Flexibility	**.09**	.01	.02	-.01	.05	.02	.04	.00	-.02
Achievement	**-.06**	.01	.01	**.05**	**.13**	-.01	.04	-.04	.00
Fitness Motivation	.02	**.10**	**.10**	**.26**	**.38**	**.06**	**.17**	**.07**	**-.09**
Interpersonal Skills - Diplomacy	-.01	.04	.02	**.06**	**.10**	.03	**.07**	-.03	.02
Stress Tolerance	**.08**	**.07**	**.07**	**.07**	**.08**	.04	**.07**	.04	-.03
Hostility to Authority	**-.13**	**-.05**	**-.10**	-.05	-.02	**-.09**	-.02	**-.10**	**.05**

Table C.1. (Continued)

Predictor Measure/Scale	Criterion Measure/Scale								
	MOS-SPEC JKT	MOS-SPEC PRS	EFFORT PRS	PHYS FIT PRS	APFT SCORE	PEERS PRS	LEADER PRS	PER DISC PRS	DISC INC
RBI (continued)									
Self-Efficacy	**.05**	.04	.03	**.06**	**.11**	.01	**.07**	-.02	-.04
Cultural Tolerance	.03	.04	.04	.01	.03	**.05**	.03	.02	-.01
Internal Locus of Control	**.12**	**.07**	**.09**	**.10**	**.08**	**.08**	**.09**	**.07**	**-.06**
Army Affective Commitment	**.09**	.01	.03	.03	.05	-.01	.03	.00	**-.08**
Respect for Authority	.00	-.00	.03	.02	.01	.03	.04	.01	.03
Narcissism	**-.06**	-.02	-.04	.03	.02	-.04	.00	**-.07**	**.06**
Gratitude	**.11**	.03	**.07**	.04	.01	**.05**	.04	**.06**	-.04
PSJT	**.23**	**.08**	**.12**	.04	.00	**.09**	**.08**	**.10**	**-.08**
AKA									
Realistic Interests	**.08**	.00	.03	.02	.00	.03	.02	.01	.00
Investigative Interests	**-.10**	-.03	**-.06**	-.04	-.02	**-.05**	-.04	**-.06**	**.05**
Artistic Interests	**-.17**	**-.06**	**-.08**	**-.05**	-.02	**-.08**	**-.06**	**-.08**	.03
Social Interests	.04	-.02	-.02	-.03	-.02	-.01	-.02	-.03	.03
Enterprising Interests	**.05**	.02	.04	.01	.02	.03	.02	.00	.00
Conventional Interests	**.10**	-.01	**.05**	.01	.01	.02	.00	.02	.00
WPA									
Realistic Interests (D)	.02	.01	-.03	.03	.01	**-.05**	.01	-.03	-.01
Mechanical (F)	.02	.01	-.03	.00	-.03	**-.05**	.02	-.03	.00
Physical (F)	.01	.02	.00	**.06**	**.08**	-.02	.01	.01	-.03
Investigative Interests (D)	-.03	.01	-.01	.00	**.05**	-.01	.01	-.02	.01
Critical Thinking (F)	**.05**	**.05**	.03	.03	**.07**	.02	.04	.01	-.01
Conduct Research (F)	**-.09**	**-.03**	-.04	-.04	.01	-.03	-.01	**-.05**	.02
Artistic Interests (D)	**-.11**	**-.04**	**-.05**	**-.04**	-.01	-.02	-.02	-.03	.02
Artistic Activities (F)	**-.13**	**-.05**	**-.07**	**-.05**	-.03	-.03	-.04	**-.04**	.04
Creativity (F)	-.03	-.01	.01	-.01	.03	.01	.02	.00	-.02
Social Interests (D)	**-.13**	-.00	-.01	.01	.04	.02	.00	-.01	.02
Work with Others (F)	**-.13**	-.01	-.02	.01	.02	.01	-.01	-.01	.01
Help Others (F)	**-.11**	.01	.01	.00	**.05**	.02	.02	-.01	.02

Table C.1. (Continued)

Predictor Measure/Scale	Criterion Measure/Scale								
	MOS-SPEC JKT	MOS-SPEC PRS	EFFORT PRS	PHYS FIT PRS	APFT SCORE	PEERS PRS	LEADER PRS	PER DISC PRS	DISC INC
WPA (continued)									
Enterprising Interests (D)	**-.12**	-.01	-.04	.02	**.05**	**-.05**	.01	**-.06**	.02
Prestige (F)	.00	.02	-.00	.03	.04	-.03	.01	-.02	.00
Lead Others (F)	**-.09**	-.01	-.04	.02	**.06**	**-.04**	.01	**-.07**	.03
High Profile (F)	**-.17**	-.03	**-.05**	.01	.03	**-.04**	.01	**-.05**	.03
Conventional Interests (D)	**-.14**	-.02	-.04	.00	.00	-.02	.00	-.04	.04
Information Management (F)	**-.17**	**-.05**	**-.06**	-.02	-.01	-.04	-.01	**-.05**	**.06**
Detail Orientation (F)	-.03	.03	.02	.02	**.04**	.02	.03	.01	.01
Clear Procedures (F)	**-.05**	.02	.01	.02	.01	.02	.01	.01	.00

Note. AFQT n = 2,085 – 2,270. AO n = 1,906 – 2,080. TAPAS-95S n = 781 – 846. AIM n = 642 – 705. RBI n = 1,638 – 1,764. PSJT n = 1,308 – 1,423. AKA n = 2,000 – 2,176. WPA n = 1,975 – 2,142. Statistically significant correlations are bolded, $p < .05$ (two-tailed). MOS-SPEC JKT = MOS-Specific Job Knowledge Test, MOS-SPEC PRS = MOS-Specific PRS Composite, EFFORT PRS = Effort PRS Composite, PHYS FIT PRS = Physical Fitness & Military Bearing PRS Composite, APFT SCORE = Last Army Physical Fitness Test (APFT) Score, PEERS PRS = Support for Peers PRS Composite, LEADER PRS = Peer Leadership PRS Composite, PER DISC PRS = Personal Discipline PRS Composite, DISC INC = Disciplinary Incidence (0 = *None*, 1 = *One or more*). D= Dimension, F = Facet.

Table C.2. Correlations between Predictor Scale Scores and Selected Retention-Related Criterion Measures

	Criterion Measure/Scale				
Predictor Measure/Scale	AFFECT COMMIT	ARMY FIT	CAR INTENT	ATTRIT COG	6-MO ATTRIT
AFQT	**-.07**	-.01	**-.05**	**.05**	**-.03**
Assembling Objects (AO)	-.02	.02	**-.05**	.03	**-.07**
TAPAS-95s					
Achievement	**.13**	**.14**	**.12**	**.13**	**-.06**
Curiosity	.03	.02	**.12**	.06	**-.04**
Non-Delinquency	.03	.03	.00	.03	-.01
Dominance	**.09**	**.07**	**.11**	.05	**.04**
Even-Temper	.04	**.07**	.07	**.11**	**-.05**
Attention-Seeking	.00	-.01	.00	-.04	**.04**
Intellectual Efficiency	.02	.04	**.07**	**.08**	-.01
Order	-.02	-.03	-.03	-.04	.00
Physical Conditioning	.06	**.12**	.05	.07	**-.10**
Tolerance	.03	.06	**.13**	.07	-.01
Cooperation/Trust	**-.14**	**-.14**	**-.16**	**-.10**	-.02
Optimism	**.09**	**.12**	**.11**	**.09**	**-.06**
AIM					
Adjustment	**.16**	**.20**	**.14**	**.17**	**-.10**
Agreeableness	**.10**	**.13**	.03	**.13**	**-.09**
Dependability	**.15**	**.15**	**.11**	**.14**	**-.07**
Leadership	**.15**	**.17**	**.17**	**.14**	**-.04**
Physical Conditioning	**.16**	**.20**	**.14**	**.14**	**-.11**
Work Orientation	**.20**	**.20**	**.22**	**.13**	**-.07**
RBI					
Peer Leadership	**.18**	**.22**	**.12**	**.16**	-.03
Cognitive Flexibility	**.10**	**.16**	**.09**	**.12**	**-.05**
Achievement	**.21**	**.22**	**.10**	**.16**	**-.05**
Fitness Motivation	**.11**	**.20**	**.08**	**.13**	**-.10**
Interpersonal Skills - Diplomacy	**.15**	**.19**	**.07**	**.14**	**-.06**
Stress Tolerance	.04	**.12**	**.05**	**.14**	**-.07**
Hostility to Authority	-.02	**-.09**	.04	**-.09**	**.04**
Self-Efficacy	**.20**	**.25**	**.14**	**.20**	**-.09**
Cultural Tolerance	**.09**	**.15**	.03	**.11**	**-.05**
Internal Locus of Control	**.15**	**.22**	.05	**.19**	**-.06**
Army Affective Commitment	**.38**	**.33**	**.28**	**.26**	**-.11**
Respect for Authority	**.17**	**.17**	**.06**	**.11**	**-.04**
Narcissism	**.14**	**.11**	**.06**	.04	.00
Gratitude	**.12**	**.13**	.01	**.12**	**-.04**
PSJT	.02	**.06**	-.03	**.05**	**-.06**

Table C.2. (Continued)

Predictor Measure/Scale	Criterion Measure/Scale				
	AFFECT COMMIT	ARMY FIT	CAR INTENT	ATTRIT COG	6-MO ATTRIT
AKA					
Realistic Interests	**.16**	**.17**	**.11**	**.14**	-.05
Investigative Interests	**.08**	**.07**	**.06**	.03	-.01
Artistic Interests	**.11**	**.07**	**.11**	.03	-.01
Social Interests	**.11**	**.11**	**.09**	**.08**	**-.04**
Enterprising Interests	**.12**	**.11**	**.09**	**.09**	-.03
Conventional Interests	**.10**	**.11**	**.07**	**.09**	**-.05**
WPA					
Realistic Interests (D)	**.19**	**.15**	**.16**	**.11**	**-.06**
Mechanical (F)	**.10**	.04	**.10**	.04	**-.05**
Physical (F)	**.20**	**.19**	**.16**	**.14**	**-.06**
Investigative Interests (D)	**.08**	**.11**	**.11**	**.09**	**-.04**
Critical Thinking (F)	**.13**	**.15**	**.12**	**.14**	**-.04**
Conduct Research (F)	.02	.04	**.06**	.02	-.02
Artistic Interests (D)	.02	.00	.03	-.01	.01
Artistic Activities (F)	.00	-.02	.02	-.03	.02
Creativity (F)	**.06**	**.05**	**.05**	**.05**	-.01
Social Interests (D)	**.12**	**.16**	**.07**	**.09**	**-.05**
Work with Others (F)	**.15**	**.17**	**.08**	**.10**	**-.07**
Help Others (F)	**.07**	**.11**	**.06**	**.06**	-.01
Enterprising Interests (D)	**.16**	**.16**	**.11**	**.09**	**-.04**
Prestige (F)	**.15**	**.16**	**.07**	**.09**	**-.05**
Lead Others (F)	**.19**	**.19**	**.15**	**.13**	**-.03**
High Profile (F)	.04	.04	**.05**	.01	-.01
Conventional Interests (D)	**.12**	**.14**	**.11**	**.08**	**-.05**
Information Management (F)	.04	**.06**	**.05**	.01	-.03
Detail Orientation (F)	**.15**	**.17**	**.14**	**.15**	**-.06**
Clear Procedures (F)	**.15**	**.17**	**.12**	**.12**	**-.06**

Note. AFQT n = 2,186-4,463. AO n = 1,997-4,170. TAPAS-95s n = 815-2,395. AIM n = 653-2,490. RBI n = 1,711-3,453. PSJT n = 1,380-1,619. AKA n = 2,091-4,153. WPA n = 2,072-4,110. Statistically significant correlations are bolded, $p <$ .05 (two-tailed). AFFECT COMMIT = Affective Commitment, ARMY FIT = Army Fit, CAR INTENT = Career Intentions, ATTRIT COG = Attrition Cognitions, 6-MO ATTRIT = 6-Month Attrition. D = Dimension, F = Facet.

Table C.3. Correlations between the AFQT and Scale Scores from the Experimental Predictor Measures

Predictor Measure/Scale	*n*	AFQT
AO	9,875	**.41**
TAPAS-95s		
Achievement	4,606	**.05**
Curiosity	4,606	**.23**
Non-Delinquency	4,606	**.06**
Dominance	4,606	**.06**
Even-Temper	4,606	**.12**
Attention-Seeking	4,606	**-.06**
Intellectual Efficiency	4,606	**.37**
Order	4,606	**-.03**
Physical Conditioning	4,606	-.02
Tolerance	4,606	.03
Cooperation/Trust	4,606	-.02
Optimism	4,606	**.18**
AIM		
Adjustment	4,775	**.11**
Agreeableness	4,669	**.09**
Dependability	4,740	**.11**
Leadership	4,787	**.12**
Physical Conditioning	4,731	.02
Work Orientation	4,722	.01
RBI		
Peer Leadership	8,567	**.13**
Cognitive Flexibility	8,567	**.25**
Achievement	8,567	**.05**
Fitness Motivation	8,567	.02
Interpersonal Skills - Diplomacy	8,567	**.04**
Stress Tolerance	8,567	**.16**
Hostility to Authority	8,567	**-.17**
Self-Efficacy	8,567	**.06**
Cultural Tolerance	8,567	**.09**
Internal Locus of Control	8,566	**.17**
Army Affective Commitment	8,567	.01
Respect for Authority	8,566	**-.04**
Narcissism	8,567	**-.06**
Gratitude	8,567	**.13**
AKA		
Realistic Interests	10,004	**.06**
Investigative Interests	10,002	**-.16**
Artistic Interests	10,003	**-.31**
Social Interests	10,004	.00
Enterprising Interests	10,005	**.02**
Conventional Interests	9,977	**.14**

Table C.3. (Continued)

Predictor Measure/Scale	*n*	AFQT
WPA		
Realistic Interests (D)	9,855	**-.13**
Mechanical (F)	9,855	**-.11**
Physical (F)	9,855	**-.10**
Investigative Interests (D)	9,855	**.07**
Critical Thinking (F)	9,854	**.14**
Conduct Research (F)	9,855	-.01
Artistic Interests (D)	9,855	**-.06**
Artistic Activities (F)	9,854	**-.11**
Creativity (F)	9,853	**.04**
Social Interests (D)	9,855	**-.12**
Work with Others (F)	9,855	**-.14**
Help Others (F)	9,855	**-.09**
Enterprising Interests (D)	9,855	**-.05**
Prestige (F)	9,855	**.04**
Lead Others (F)	9,853	**-.06**
High Profile (F)	9,855	**-.10**
Conventional Interests (D)	9,855	**-.19**
Information Management (F)	9,855	**-.18**
Detail Orientation (F)	9,855	**-.08**
Clear Procedures (F)	9,855	**-.13**

Note. Statistically significant correlations are bolded, $p < .05$ (two-tailed).

Table C.4. Correlations between Scales Scores from the TAPAS-95s and Other Temperament Predictor Measures

	TAPAS-95s Scale											
Measure/Scale	ACH	CUR	DEL	DOM	TEM	ATT	INT	ORD	PHY	TOL	TRU	OPT
AIM												
Adjustment	**.12**	**.20**	**.16**	**.05**	**.32**	**-.18**	**.13**	.00	**.09**	**.12**	-.03	**.38**
Agreeableness	**.08**	**.17**	**.26**	**-.04**	**.40**	**-.26**	**.05**	.00	**.05**	**.07**	**.07**	**.19**
Dependability	**.16**	**.16**	**.46**	**.10**	**.15**	**-.32**	**.08**	**.11**	-.01	**.06**	-.02	**.07**
Leadership	**.19**	**.22**	.03	**.49**	.02	**.05**	**.23**	**.05**	**.09**	**.13**	**-.23**	**.06**
Physical Conditioning	**.22**	**.10**	.00	**.04**	**.06**	**-.06**	.03	**.05**	**.60**	**.06**	**-.12**	**.04**
Work Orientation	**.36**	**.23**	**.05**	**.21**	**.12**	**-.08**	**.17**	**.09**	**.30**	**.12**	**-.23**	**.08**
RBI												
Peer Leadership	**.15**	**.21**	.02	**.41**	**.04**	**.08**	**.23**	.03	**.13**	**.17**	**-.19**	**.06**
Cognitive Flexibility	**.13**	**.41**	**.09**	**.17**	**.17**	**-.09**	**.33**	-.03	.03	**.25**	**-.09**	**.08**
Achievement	**.23**	**.19**	**.18**	**.22**	.01	**-.06**	**.14**	**.11**	**.13**	**.13**	**-.13**	-.03
Fitness Motivation	**.17**	**.06**	**-.10**	**.08**	**.04**	.01	**.06**	-.02	**.61**	.01	**-.18**	**.08**
Interpersonal Skills - Diplomacy	**.09**	**.15**	.01	**.30**	**.06**	**.17**	**.11**	.02	**.10**	**.16**	**-.07**	**.10**
Stress Tolerance	**.16**	**.17**	**.05**	**.08**	**.25**	**-.12**	**.20**	-.01	**.14**	**.09**	**-.07**	**.30**
Hostility to Authority	**-.14**	**-.18**	**-.44**	**-.06**	**-.18**	**.34**	**-.09**	**-.08**	**.05**	**-.08**	**-.07**	**-.10**
Self-Efficacy	**.24**	**.20**	**.06**	**.24**	**.12**	-.03	**.19**	**.06**	**.19**	**.15**	**-.17**	**.16**
Cultural Tolerance	**.12**	**.23**	**.17**	**.16**	**.18**	**-.10**	**.16**	.01	-.01	**.35**	-.02	**.12**
Internal Locus of Control	**.21**	**.18**	**.14**	**.15**	**.15**	**-.08**	**.16**	**.08**	**.10**	**.11**	**-.05**	**.21**
Army Affective Commitment	**.19**	**.10**	**.10**	**.12**	**.10**	**-.07**	.03	.02	**.14**	**.11**	**-.11**	**.14**
Respect for Authority	**.15**	**.07**	**.18**	**.07**	**.02**	**-.06**	.00	**.04**	.00	**.07**	-.02	**-.05**
Narcissism	**.06**	**.07**	**-.08**	**.20**	**-.10**	**.12**	**.09**	**.07**	**.12**	**.08**	**-.16**	.00
Gratitude	**.11**	**.14**	**.18**	**.11**	**.10**	**-.05**	**.05**	**.05**	.01	**.09**	.02	**.06**
PSJT	**.15**	**.16**	**.36**	**.11**	**.17**	**-.23**	.02	.07	-.02	**.10**	.07	**.10**

Note. AIM n = 3,685-3,723. RBI n = 3,426. PSJT n = 523. Statistically significant correlations are bolded, $p < .05$ (two-tailed). ACH = Achievement, CUR = Curiosity, DEL = Non-Delinquency, DOM = Dominance, TEM = Even-Temper, ATT = Attention-Seeking, INT = Intellectual Efficiency, ORD = Order, PHY = Physical Conditioning, TOL = Tolerance, TRU = Cooperation/Trust, OPT = Optimism.

Table C.5. Correlations between Scale Scores from the WPA and the AKA

	AKA Scale					
WPA Scale	REAL	INVEST	ART	SOC	ENTER	CONV
Realistic Interests (D)	**.15**	**.13**	**.15**	**.11**	**.08**	**.05**
Mechanical (F)	**.08**	**.10**	**.14**	**.05**	**.04**	.00
Physical (F)	**.18**	**.12**	**.11**	**.13**	**.10**	**.08**
Investigative Interests (D)	**.21**	**.20**	**.13**	**.20**	**.20**	**.20**
Critical Thinking (F)	**.26**	**.17**	**.06**	**.21**	**.23**	**.25**
Conduct Research (F)	**.10**	**.18**	**.16**	**.13**	**.13**	**.11**
Artistic Interests (D)	**.05**	**.16**	**.18**	**.09**	**.10**	**.06**
Artistic Activities (F)	.00	**.14**	**.18**	**.06**	**.06**	.01
Creativity (F)	**.14**	**.14**	**.10**	**.13**	**.15**	**.13**
Social Interests (D)	**.23**	**.22**	**.16**	**.26**	**.21**	**.20**
Work with Others (F)	**.24**	**.20**	**.14**	**.25**	**.20**	**.20**
Help Others (F)	**.17**	**.19**	**.14**	**.21**	**.17**	**.16**
Enterprising Interests (D)	**.20**	**.20**	**.15**	**.19**	**.20**	**.18**
Prestige (F)	**.25**	**.16**	**.06**	**.20**	**.20**	**.22**
Lead Others (F)	**.21**	**.18**	**.12**	**.18**	**.18**	**.17**
High Profile (F)	**.03**	**.14**	**.17**	**.07**	**.09**	**.05**
Conventional Interests (D)	**.19**	**.26**	**.24**	**.21**	**.20**	**.16**
Information Management (F)	**.06**	**.21**	**.23**	**.12**	**.12**	**.08**
Detail Orientation (F)	**.25**	**.19**	**.13**	**.21**	**.19**	**.19**
Clear Procedures (F)	**.25**	**.20**	**.15**	**.21**	**.20**	**.19**

Note. n = 9,593 – 9,618. Statistically significant correlations are bolded, $p < .05$ (two-tailed). REAL = Realistic Interests, INVEST = Investigative Interests, ART = Artistic Interests, SOC = Social Interests, ENTER = Enterprising Interests, CONV = Conventional Interests.

Table C.6. Correlations between Scale Scores from the TAPAS-95s and the WPA

	TAPAS-95s Scale											
WPA Scale	ACH	CUR	DEL	DOM	TEM	ATT	INT	ORD	PHY	TOL	TRU	OPT
Realistic Interests (D)	**.13**	.00	**-.08**	**-.04**	**.03**	.00	**-.07**	**-.04**	**.24**	**-.04**	**-.10**	**.05**
Mechanical (F)	**.10**	.03	**-.10**	**-.10**	.01	-.02	**-.04**	-.02	**.08**	**-.07**	**-.05**	.03
Physical (F)	**.13**	-.02	**-.05**	.02	**.04**	.02	**-.06**	**-.04**	**.34**	.00	**-.11**	**.05**
Investigative Interests (D)	**.16**	**.39**	**.10**	**.12**	**.15**	**-.12**	**.25**	**.04**	.02	**.17**	**-.09**	**.03**
Critical Thinking (F)	**.21**	**.33**	**.11**	**.18**	**.15**	**-.12**	**.26**	**.05**	**.06**	**.15**	**-.12**	**.08**
Conduct Research (F)	**.08**	**.35**	**.06**	**.05**	**.10**	**-.10**	**.17**	.01	-.02	**.15**	**-.05**	-.02
Artistic Interests (D)	**-.04**	**.17**	**-.07**	.02	**.05**	.02	**.03**	**-.03**	**-.03**	**.14**	.02	**-.05**
Artistic Activities (F)	**-.08**	**.10**	**-.07**	**-.04**	.02	.03	-.03	**-.04**	**-.04**	**.11**	**.04**	**-.07**
Creativity (F)	**.05**	**.25**	**-.04**	**.12**	**.10**	.00	**.14**	-.01	-.01	**.14**	-.03	.01
Social Interests (D)	**.05**	**.11**	**.15**	**.19**	**.08**	**-.04**	**-.05**	.02	.02	**.16**	-.01	**-.07**
Work with Others (F)	**.06**	**.09**	**.11**	**.16**	**.08**	-.01	**-.06**	.01	**.06**	**.15**	.00	-.03
Help Others (F)	**.04**	**.11**	**.16**	**.18**	**.05**	**-.06**	**-.03**	.03	-.02	**.14**	-.01	**-.09**
Enterprising Interests (D)	**.10**	**.14**	.00	**.27**	.02	**.07**	**.05**	**.05**	**.08**	**.12**	**-.13**	**-.06**
Prestige (F)	**.13**	**.15**	**.08**	**.22**	**.04**	.01	**.08**	**.09**	**.08**	**.10**	**-.10**	-.01
Lead Others (F)	**.10**	**.10**	-.02	**.36**	.00	**.08**	**.04**	.03	**.10**	**.11**	**-.14**	-.03
High Profile (F)	.00	**.08**	**-.06**	**.09**	.00	**.06**	.00	.01	.01	**.08**	**-.06**	**-.10**
Conventional Interests (D)	**.12**	**.09**	**.16**	**.08**	**.03**	**-.10**	-.01	**.15**	-.02	**.07**	**-.04**	**-.07**
Information Management (F)	**.04**	**.06**	**.06**	**.05**	.00	**-.05**	-.02	**.08**	**-.06**	**.06**	-.01	**-.10**
Detail Orientation (F)	**.19**	**.16**	**.15**	**.10**	**.07**	**-.13**	**.09**	**.16**	**.04**	**.08**	**-.08**	.02
Clear Procedures (F)	**.15**	**.11**	**.20**	**.08**	**.06**	**-.14**	.01	**.18**	.02	**.06**	**-.06**	-.01

Note. n = 4,343. Statistically significant correlations are bolded, $p < .05$ (two-tailed). ACH = Achievement, CUR = Curiosity, DEL = Non-Delinquency, DOM = Dominance, TEM = Even-Temper, ATT = Attention-Seeking, INT = Intellectual Efficiency, ORD = Order, PHY = Physical Conditioning, TOL = Tolerance, TRU = Cooperation/Trust, OPT = Optimism.

Table C.7. Intercorrelations among Scale Scores from Selected Performance-Related Criterion Measures

Scale		1	2	3	4	5	6	7	8
1	MOS-Specific Job Knowledge Test								
2	MOS-Specific PRS Composite	**.15**							
3	Effort PRS Composite	**.20**	**.63**						
4	Physical Fitness & Bearing PRS Composite	**.10**	**.60**	**.74**					
5	Last APFT Score	.00	**.23**	**.25**	**.44**				
6	Support for Peers PRS Composite	**.15**	**.60**	**.74**	**.63**	**.13**			
7	Peer Leadership PRS Composite	**.12**	**.64**	**.76**	**.70**	**.27**	**.73**		
8	Personal Discipline PRS Composite	**.18**	**.58**	**.79**	**.64**	**.15**	**.80**	**.69**	
9	Disciplinary Incidence	**-.12**	**-.20**	**-.26**	**-.21**	**-.13**	**-.17**	**-.23**	**-.29**

Note. n = 2,030 – 2,277. APFT = Army Physical Fitness Test. Disciplinary Incidence is a constructed variable based on the self-reported number of disciplinary incidents and is coded 0 = *None* and 1 = *One or more*. Statistically significant correlations are bolded, $p < .05$ (two-tailed).

Table C.8. Intercorrelations among Scale Scores from Selected Retention-Related Criterion Measures

Scale		1	2	3	4
1	Affective Commitment				
2	Army Fit	.79			
3	Career Intentions	.58	.55		
4	Attrition Cognitions	.65	.72	.53	
5	6-Month Attrition	-.12	-.15	-.08	-.21

Note. n = 1,041 – 2,178. All correlations are statistically significant, $p < .05$ (two-tailed).

APPENDIX D
PREDICTOR SCORE SUBGROUP DIFFERENCES

Table D.1. Standardized Mean Differences (Cohen's d) by Subgroup Combination and Predictor Measure

	Gender Differences					Race Differences					Ethnicity Differences				
	Female (F)		Male (M)		F-M	Black (B)		White (W)		B-W	Hispanic (H)		White, Non-Hispanic (WNH)		H-WNH
Predictor	*M*	*SD*	*M*	*SD*	*d*	*M*	*SD*	*M*	*SD*	*d*	*M*	*SD*	*M*	*SD*	*d*
AFQT	53.48	18.17	56.77	19.53	-0.17	47.15	16.46	57.95	19.29	-0.56	51.24	17.75	58.95	19.31	-0.40
AO	53.71	7.73	55.16	7.98	-0.18	51.03	8.54	55.54	7.62	-0.59	54.75	7.76	55.60	7.64	-0.11
AIM															
Adjustment	1.24	0.32	1.26	0.29	-0.07	1.29	0.25	1.26	0.30	0.11	1.29	0.27	1.25	0.30	0.14
Agreeableness	1.27	0.29	1.26	0.26	0.06	1.29	0.25	1.25	0.27	0.15	1.29	0.25	1.25	0.27	0.15
Dependability	1.34	0.27	1.25	0.28	0.32	1.31	0.27	1.25	0.28	0.20	1.28	0.27	1.25	0.29	0.09
Leadership	1.25	0.29	1.20	0.28	0.19	1.24	0.26	1.20	0.28	0.14	1.21	0.26	1.20	0.29	0.04
Physical Conditioning	1.13	0.33	1.21	0.34	-0.22	1.22	0.30	1.19	0.34	0.09	1.22	0.31	1.19	0.35	0.07
Work Orientation	1.23	0.29	1.20	0.29	0.09	1.25	0.26	1.20	0.29	0.19	1.22	0.28	1.19	0.30	0.08
Average Absolute d					*0.16*					*0.15*					*0.10*
TAPAS-95s															
Achievement	0.27	0.64	0.15	0.64	0.19	0.13	0.58	0.19	0.65	-0.09	0.14	0.65	0.19	0.65	-0.07
Curiosity	-0.02	0.82	-0.09	0.79	0.08	-0.02	0.75	-0.09	0.81	0.09	-0.04	0.79	-0.10	0.81	0.07
Non-Delinquency	0.30	0.64	0.04	0.64	0.41	0.07	0.61	0.09	0.65	-0.03	0.03	0.61	0.10	0.66	-0.10
Dominance	-0.03	0.60	-0.18	0.61	0.24	-0.05	0.58	-0.16	0.61	0.18	-0.17	0.60	-0.16	0.61	-0.01
Even-Temper	-0.55	0.81	-0.44	0.75	-0.14	-0.45	0.72	-0.46	0.77	0.02	-0.46	0.77	-0.46	0.77	0.01
Attention-Seeking	-0.15	0.79	-0.13	0.79	-0.02	-0.14	0.82	-0.13	0.79	-0.02	-0.14	0.79	-0.13	0.79	-0.01
Intellectual Efficiency	-0.26	0.64	-0.17	0.64	-0.14	-0.15	0.60	-0.19	0.64	0.07	-0.25	0.60	-0.18	0.65	-0.11
Order	0.13	0.62	-0.07	0.63	0.31	0.02	0.61	-0.04	0.64	0.09	-0.01	0.62	-0.05	0.64	0.06
Physical Conditioning	-0.04	0.73	0.16	0.70	-0.29	0.15	0.68	0.12	0.71	0.04	0.16	0.73	0.12	0.71	0.06
Tolerance	-0.28	0.62	-0.46	0.68	0.27	-0.26	0.66	-0.46	0.67	0.29	-0.34	0.62	-0.47	0.68	0.19
Cooperation/Trust	-0.21	0.84	-0.32	0.86	0.14	-0.32	0.87	-0.30	0.86	-0.02	-0.29	0.85	-0.31	0.86	0.01
Optimism	-0.12	0.61	-0.06	0.59	-0.11	-0.10	0.57	-0.07	0.60	-0.05	-0.07	0.61	-0.07	0.59	0.00
Average Absolute d					*0.20*					*0.08*					*0.06*
PSJT	4.79	0.36	4.64	0.42	0.36	4.60	0.45	4.69	0.40	-0.21	4.66	0.40	4.69	0.40	-0.08

Table D.1. (Continued)

	Gender Differences					Race Differences					Ethnicity Differences				
	Female (F)		Male (M)		F-M	Black (B)		White (W)		B-W	Hispanic (H)		White, Non-Hispanic (WNH)		H-WNH
Predictor	*M*	*SD*	*M*	*SD*	*d*	*M*	*SD*	*M*	*SD*	*d*	*M*	*SD*	*M*	*SD*	*d*
RBI															
Peer Leadership	3.73	0.64	3.57	0.64	0.26	3.68	0.67	3.59	0.64	0.15	3.59	0.65	3.59	0.64	-0.01
Cognitive Flexibility	3.58	0.62	3.45	0.64	0.20	3.57	0.63	3.45	0.64	0.19	3.52	0.64	3.44	0.64	0.12
Achievement	3.75	0.56	3.49	0.57	0.45	3.74	0.59	3.51	0.57	0.42	3.57	0.58	3.50	0.56	0.12
Fitness Motivation	2.93	0.65	3.40	0.65	-0.72	3.28	0.71	3.31	0.67	-0.03	3.32	0.67	3.30	0.68	0.03
Interpersonal Skills - Diplomacy	3.85	0.74	3.60	0.74	0.33	3.77	0.73	3.64	0.75	0.18	3.69	0.75	3.63	0.75	0.08
Stress Tolerance	2.93	0.54	3.03	0.50	-0.19	3.03	0.53	3.01	0.51	0.04	3.02	0.52	3.00	0.51	0.03
Hostility to Authority	2.25	0.59	2.59	0.65	-0.52	2.50	0.68	2.52	0.65	-0.03	2.52	0.67	2.52	0.65	0.00
Self-Efficacy	4.11	0.59	3.99	0.63	0.19	4.17	0.60	3.99	0.62	0.27	4.05	0.62	3.99	0.62	0.09
Cultural Tolerance	3.97	0.65	3.70	0.74	0.38	3.89	0.70	3.72	0.73	0.23	3.96	0.71	3.69	0.73	0.38
Internal Locus of Control	3.66	0.55	3.52	0.58	0.25	3.58	0.57	3.54	0.58	0.06	3.54	0.57	3.54	0.58	0.00
Army Affective Commitment	3.68	0.71	3.74	0.68	-0.09	3.59	0.69	3.76	0.68	-0.24	3.76	0.67	3.75	0.68	0.01
Respect for Authority	3.67	0.67	3.47	0.69	0.29	3.60	0.75	3.49	0.68	0.16	3.52	0.69	3.49	0.68	0.04
Narcissism	3.63	0.56	3.60	0.57	0.05	3.84	0.59	3.57	0.55	0.50	3.67	0.57	3.56	0.55	0.21
Gratitude	4.11	0.66	3.90	0.72	0.29	3.87	0.77	3.96	0.70	-0.14	3.93	0.73	3.96	0.70	-0.06
Average Absolute d					*0.30*					*0.19*					*0.08*
AKA															
Realistic	4.09	0.58	4.04	0.61	0.09	4.07	0.63	4.05	0.60	0.04	4.04	0.63	4.05	0.60	-0.01
Investigative	3.43	0.74	3.37	0.73	0.08	3.50	0.72	3.36	0.73	0.19	3.41	0.74	3.36	0.73	0.07
Artistic	2.72	0.96	2.76	0.92	-0.04	2.99	0.93	2.70	0.92	0.31	2.82	0.92	2.69	0.92	0.14
Social	3.87	0.70	3.76	0.71	0.15	3.85	0.74	3.77	0.70	0.11	3.78	0.73	3.77	0.70	0.03
Enterprising	3.80	0.70	3.67	0.71	0.18	3.78	0.72	3.68	0.70	0.14	3.68	0.71	3.68	0.70	0.00
Conventional	4.02	0.68	3.91	0.69	0.16	3.95	0.70	3.93	0.69	0.04	3.90	0.71	3.93	0.69	-0.06
Average Absolute d					*0.12*					*0.14*					*0.05*

Table D.1. (Continued)

	Gender Differences					Race Differences					Ethnicity Differences				
	Female (F)		Male (M)		F-M	Black (B)		White (W)		B-W	Hispanic (H)		White, Non-Hispanic (WNH)		H-WNH
Predictor	*M*	*SD*	*M*	*SD*	*d*	*M*	*SD*	*M*	*SD*	*d*	*M*	*SD*	*M*	*SD*	*d*
WPA Dimensions															
Realistic (R)	3.00	0.84	3.63	0.73	-0.87	3.17	0.90	3.56	0.76	-0.52	3.52	0.81	3.56	0.75	-0.05
Investigative (I)	3.32	0.67	3.27	0.65	0.08	3.40	0.67	3.24	0.64	0.24	3.40	0.68	3.22	0.63	0.27
Artistic (A)	2.87	0.82	2.77	0.74	0.14	2.95	0.79	2.75	0.75	0.26	2.90	0.79	2.73	0.74	0.22
Social (S)	3.87	0.64	3.53	0.64	0.55	3.81	0.65	3.55	0.64	0.40	3.72	0.64	3.53	0.64	0.29
Enterprising (E)	3.36	0.60	3.36	0.59	0.00	3.56	0.63	3.32	0.57	0.41	3.48	0.60	3.30	0.57	0.31
Conventional (C)	3.40	0.67	3.19	0.60	0.36	3.54	0.67	3.17	0.60	0.63	3.37	0.64	3.14	0.58	0.40
Average Absolute d					*0.33*					*0.41*					*0.26*
WPA Facets															
Mechanical (R)	2.55	1.05	3.37	0.99	-0.83	2.92	1.12	3.25	1.03	-0.31	3.22	1.04	3.24	1.04	-0.02
Physical (R)	3.35	0.92	3.83	0.80	-0.60	3.36	0.96	3.80	0.81	-0.55	3.76	0.87	3.80	0.80	-0.05
Critical Thinking (I)	3.80	0.73	3.75	0.72	0.06	3.83	0.73	3.75	0.72	0.12	3.82	0.74	3.74	0.71	0.12
Conduct Research (I)	2.84	0.79	2.78	0.77	0.08	2.97	0.80	2.75	0.76	0.29	2.97	0.81	2.71	0.75	0.34
Artistic Activities (A)	2.48	0.91	2.37	0.85	0.14	2.55	0.90	2.35	0.85	0.24	2.51	0.90	2.33	0.84	0.21
Creativity (A)	3.64	0.89	3.57	0.85	0.08	3.73	0.90	3.56	0.85	0.20	3.68	0.87	3.54	0.85	0.16
Work with Others (S)	3.97	0.69	3.76	0.71	0.29	3.97	0.72	3.77	0.71	0.28	3.93	0.71	3.75	0.71	0.25
Help Others (S)	3.78	0.74	3.29	0.72	0.68	3.65	0.76	3.33	0.74	0.43	3.51	0.75	3.31	0.74	0.26
Prestige(E)	3.91	0.65	3.87	0.66	0.06	3.99	0.69	3.86	0.65	0.20	3.96	0.66	3.84	0.65	0.19
Lead Others (E)	3.56	0.77	3.56	0.74	0.00	3.69	0.79	3.53	0.73	0.22	3.68	0.76	3.51	0.72	0.23
High Profile (E)	2.48	0.90	2.53	0.87	-0.05	2.89	0.93	2.44	0.85	0.53	2.66	0.89	2.41	0.84	0.30
Information Management (C)	2.86	0.91	2.57	0.80	0.36	3.12	0.89	2.53	0.79	0.74	2.82	0.86	2.49	0.78	0.43
Detail Orientation (C)	4.00	0.79	3.85	0.77	0.19	4.02	0.80	3.85	0.78	0.21	3.99	0.79	3.83	0.77	0.21
Clear Procedures (C)	4.06	0.74	3.86	0.76	0.26	4.08	0.76	3.86	0.75	0.30	4.01	0.78	3.84	0.74	0.23
Average Absolute d					*0.26*					*0.33*					*0.21*

Note. *M* = Scale mean for group, *SD* = Scale standard deviation for group; $d = (M_{COMPARISON} - M_{REFERENT})/SD_{REFERENT}$. The referent groups are Males, Whites, and Non-Hispanic Whites; the comparison groups are Females, Blacks, and Hispanics. The WPA yields six dimension and 14 facet scores. The letters in parentheses after the name of the facet scores denotes the higher order dimension.